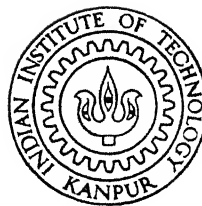# ROLE OF DETERMINISTIC CHAOS IN SPEECH SIGNAL MODELING

by

ARUN KUMAR

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

MARCH, 1990

Role of Deterministic Chaos in Speech Signal Modeling

A Thesis Submitted
in Partial Fulfilment of the Requirements
for the Degree of
Master of Technology

by

ARUN KUMAR

to the

DEPARTMENT OF ELECTRICAL ENGINEERING
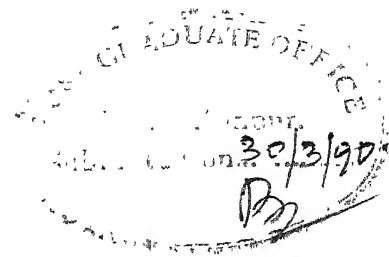
INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

March, 1990.

CERTIFICATE

It is certified that the work contained in the thesis entitled *Role of Deterministic Chaos in Speech Signal Modeling* by *Arun Kumar*, has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

S. K. Mullick
Professor
Department of Electrical Engineering
I.I.T., Kanpur.

March, 1990.

ACKNOWLEDGEMENTS

CONTENTS

CHAPTER 1

INTRODUCTION

Probabilistic and deterministic description of natural phenomena have coexisted for ages. Very often, probabilistic descriptions are assumed derivable from the underlying determinism although this is rarely carried out. This is because probabilistic descriptions are invoked when either the deterministic model equations become hard to solve exactly due to many degrees of freedom or the probabilistic models themselves are adequate for the problem at hand. Just as probabilistic descriptions are contrived conveniences, so also are deterministic descriptions in the form of differential equations, because scientific theories are not discoveries of the laws of nature but rather, inventions of the human mind. The only difference between the two descriptions then appears to be that while deterministic descriptions examine the results of a single trial and predict the future in certain terms, probabilistic descriptions are presented as statements of average behaviour and the future is predicted with a quantitative element of uncertainty.

However, there has been a distortion of facts due to blind faith or excessive use of probabilistic descriptions. Many people speak of random processes as though they are a fundamental source of randomness which is misleading. The only truly fundamental source of randomness known is the uncertainty principle. However, as a matter of course, events like roulette wheel spins, dice throws and coin tosses are presumed to be completely random. It was pointed out early in this century by Poincare that many of the so called classic examples of randomness in fact are quite deterministic and involve only a few degrees of freedom.

At this juncture, a brief discussion on randomness is in order Conventionally, randomness may be deduced from impulse-at-zero-delay autocorrelation or a flat spectrum. However, many time series which pass these conventional tests of randomness are now known to be deterministic in as much as they can be realized by simple differential or difference equations. Hence, a more fundamental notion of randomness relevant to the development of the viewpoint followed here is necessary This is based on algorithmic complexity theory developed largely by Kolmogorov and Chaitin [23,24] They define a measure of *how random* a particular number or time series is to be the length of the algorithm required to specify that number or time series If a number or time series is completely random, the only way to specify it is to write it down. Thus, the algorithm will be as long as the number or time series On the other hand, a more ordered time series can be generated by an algorithm much shorter than the actual time series.

The apparent randomness, or more appropriately, the unpredictability of the outcome of roulette wheel spins, dice throws and coin tosses comes from *sensitive dependence on initial conditions* i.e, a small perturbation causes a much larger effect at a later time. When sensitive dependence comes in a sustained way, i.e., there is no settling down as in a coin toss, it is called *chaos* (precise definitions will be given in Chapter 2). In this sense, roulette wheel spins etc. have sensitive dependence on initial conditions, but are not chaotic because they come to rest. Chaos is, thus, a special case of sensitive dependence on initial conditions, where there is also sustained motion. Chaos is defined in the context of deterministic dynamics, and therefore, in a strict sense, it is not random. Such a conflict does not arise in the domain of the above definition of randomness although chaotic trajectories may pass the other classic tests of randomness.

Ultimately then, uncertainty originates from something external to the dynamics eg. measurement error or external "noise". But the sensitive dependence exaggerates uncertainty so that small uncertainties in initial conditions turn into large ones. Since chaos amplifies noise exponentially, any uncertainty is amplified to macroscopic proportions in finite time, and short term determinism becomes long term unpredictability

This thesis is concerned with some aspects of time series modeling and the role that chaotic models can play in this venture Let us first briefly review some of the popular time series modeling schemes.

1.1 Review of some Signal Modeling Schemes

A signal or time series is a manifestation of some phenomenon under investigation Generally, signals are continuous-time in nature but the process of observation at intervals makes them discrete-time. Modeling of signals is done for various reasons, primary among them being, to *predict* the future behaviour of the phenomenon, to *filter* out the unwanted component or to *smooth* the signal to reduce external noise or measurement error that is not a part of the signal.

Signal modeling schemes can be broadly divided into two categories, namely *auto-regressive modeling* and *state-space modeling*. While the theory of auto-regressive modeling is rooted in probabilistic considerations ie. the signal is considered to be a sample function or a realization of a random process, the theory of state-space modeling has developed from both deterministic and probabilistic viewpoints. Let us now discuss these two modeling schemes. More emphasis will be given to the deterministic state-space modeling approach as it naturally extends to the case of signal modeling schemes that will be pursued in the thesis.

1. Auto-regressive modeling :- Consider a stationary random process $\{y_0, y_1, ..\}$.

An *auto-regressive* (AR) *process* is defined as one having the form

$$y_n + a_1 \ y_{n-1} + ... + a_p \cdot y_{n-p} = e_n \tag{1.1.1}$$

where $\{e_n\}$ is a white noise sequence such that

$$E [e_n \cdot e_l] = 0, \quad n \neq l$$

$$E [e_n \cdot y_l] = 0, \quad l < n \tag{1.1.2}$$

and p is the *order* of auto-regression.

Now, given a time series $\{y_0, y_1, ..\}$, the assumption that it is a realization of a stationary random process allows one to construct an auto-regressive model,

$$y_n = \sum_{k=1}^{p} a_k \cdot y_{n-k} + \xi_n \tag{1.1.3}$$

The order of auto-regression and the coefficients a_k are chosen so that the residuals ξ_n approximate a white noise sequence i.e.,

$$\xi_n = y_n - \hat{y}_n, \quad n = 0, 1, ... \tag{1.1.4}$$

where,

$$\xi_0 = y_0,$$

and,

$\hat{y}_n$ = linear least squares error estimate of y_n given $\{y_{n-p}, ...,$ $y_{n-1}\}$

$$= \sum_{k=1}^{p} a_k \cdot y_{n-k} . \tag{1.1.5}$$

To obtain the model coefficients $\{a_k\}$ then requires the solution of

$$\sum_{k=1}^{p} a_k \cdot R (|i - k|) = R(i), \quad 1 \leq i \leq p \tag{1.1.6}$$

where $R(k)$ denotes the autocorrelation function of the stationary random process $\{y_0, y_1, ...\}$.

An estimate of the autocorrelation function can be made from a finite length N of actual data using the following :

$$R(k) = \sum_{m=0}^{N-1-k} y(m) \cdot y(m+k) \qquad (1.1.7)$$

Such a scheme is also referred to as *linear prediction coding* (LPC) scheme. There are various formulations of the LPC scheme of which we shall briefly discuss the autocorrelation and the covariance methods

In the *autocorrelation method*, modeling is done using blocks of data such that values outside the block are assumed to be zero Thus, given one such block ($y_0, \ldots, y_{N-1}$), the linear least squares problem is to minimize

$$E = \sum_{n=-\infty}^{\infty} \xi_n^2 = \sum_{n=0}^{N+p+1} \xi_n^2 \qquad (1.1 8)$$

where ξ_n is given by equation (1.1 4) and is zero outside the interval (0, N+p-1). It leads to the requirement of solving equation (1.1.6) using equation (1.1 7) In this approach, the residuals ξ_n do not approach a white noise approximation because of the arbitrary assumption that the signal is zero outside the finite blocklength.

An alternative to the above approach is the *covariance method* in which the assumption is that the residual length over which the total error is minimized is fixed i.e.,

$$E = \sum_{n=0}^{N-1} \xi_n \qquad (1.1.9)$$

This leads to a set of equations slightly different from (1.1.6) and (1.1.7), namely,

$$\sum_{k=1}^{p} a_k \cdot \varphi(i-k) = \varphi(i), \qquad 1 \le i \le p \qquad (1.1.10)$$

where

$$\varphi(i-k) = \sum_{m=-k}^{N-k-1} y(m) \cdot y(m+k-1), \quad 1 \le i \le p, \quad 0 \le k \le p. \qquad (1.1.11)$$

Thus, it is seen that a block of data ($y_{-k}, \ldots, y_0, \ldots, y_{N-1}$) is required in this case. Since the data is not arbitrarily truncated to zero in this case

outside a fixed blocklength, the problem of large residuals near the ends of the block is overcome. However, equations (1.1.10) - (1.1.11) do not have the Toeplitz structure as in equations (1.1.6) - (1 1 7) Hence obtaining the parameters $\{a_k\}$ in the case of covariance method requires more computation than in the case of autocorrelation method

A moving-average scheme for a stationary random process $\{ y_0, y_1, \ldots \}$ is one where

$$y_n = \xi_n - \sum_{j=1}^{q} b_j \cdot \xi_{n-j} , \qquad (1.1.12)$$

that is y_n can be expressed in terms of the residuals ξ_n and ξ_{n-j} as given by equation (1.1.4)

Finally, it was shown by Wold in 1954 that any discrete, stationary time series $\{ y_n \}$ can be broken into ·

$$y_n = \sum_{k=1}^{p} a_k \cdot y_{n-k} - \sum_{j=1}^{q} b_j \cdot \xi_{n-j} + \sum_{r=1}^{h} \sum_{q=1}^{s} c_{rq} \cdot \cos \left(\frac{2\pi n q}{T} + c_{rq} \right)$$

$$(1.1.13)$$

For practical purposes, the last term in equation (1.1.13) is included in the error term. This then gives the *autoregressive - moving average* (ARMA) modeling scheme which can mathematically be expressed as follows :

$$y_n = \sum_{k=1}^{p} a_k \cdot y_{n-k} - \sum_{j=1}^{q} b_j \cdot \xi_{n-j} + \xi_n . \qquad (1.1.14)$$

The AR, MA and ARMA modeling schemes have been studied in great detail. Also, variations of these schemes were studied to include other features like seasonality, trend, cyclicality of data etc. The algorithms were extended to cope with non-stationary processes also. Finally, the presence of efficient algorithms to implement them has made these modeling schemes very popular in diverse fields.

2 State - Space Modeling

- This review which follows the development given in [70], shall be restricted to deterministic models only i e, the time series will be considered to be the output of a deterministic process Stochastic state-space modeling schemes have not been discussed in this review because they are not directly related to the thesis work This means that some of the very popular schemes eg Kalman filtering concepts etc have not been discussed

Generally, any natural phenomena can be modeled as a dynamical system which can then be described by a system of differential equations Three successively more general types of model equations will be considered Let $X(t)$ be a vector of functions i e ,

$$X(t) = [x_0(t) \, x_1(t) \quad x_m(t)]^T \tag{1 1 15}$$

and define its time derivative as

$$X(t) = [x_0(t) \, x_1(t) \quad x_m(t)]^T \tag{1 1 16}$$

Then the three classes of differential equations are

$$X(t) = A \, X(t) \tag{1 1 17}$$

$$X(t) = A(t) \, X(t) \tag{1 1 18}$$

$$X(t) = F[X(t) \, t] \tag{1 1 19}$$

The vector $X(t)$ is called the *state vector* of the model and its components are the *state variables* In equation (1 1 17), $X(t)$ is related to $X(t)$ by a linear transformation defined by the constant matrix A Such a system is accordingly called a *constant coefficient linear differential equation* In equation (1 1 18), $X(t)$ is a linear transformation on $X(t)$ with the transformation matrix $A(t)$ having time varying components Such a system is called a *time varying linear differential equation* In equation (1 1 19), $X(t)$ is related to $X(t)$ by a vector of functions of the form

$$x_0(t) = f_0 [x_0(t), x_1(t), \quad x_m(t), t]$$

$$x_1(t) = f_1 [x_0(t), x_1(t), \quad x_m(t), t]$$

$$x_m(t) = f_m [x_0(t), x_1(t), \quad x_m(t), t] \tag{1 1 20}$$

where f_0, f_1, , f_m are possibly non-linear functions of $X(t)$ and possibly t as well It is, thus, the general form of a non-linear differential equation

The observation of the process provides us with *vectors of observations* which ideally are linear or non-linear transformations of some or all of the state-variables In practice, the measuring instruments introduce errors which are assumed to be additive

Thus, letting X_n be the vector of model state-variables at $t=t_n$ and letting Y_n be the vector of observations obtained at that time, the three successively more general observation schemes are

$$Y_n = M X_n + N_n , \tag{1 1 21}$$

$$Y_n = M_n X_n + N_n , \tag{1 1 22}$$

$$Y_n = G [X_n, t_n] + N_n \tag{1 1.23}$$

We consider the following examples of the above ideas Let the state-variables of the model be the position and velocity , in Cartesian coordinates, of a body in straight line motion Then, its differential equation will be

$$\frac{d}{dt} \begin{bmatrix} x_0(t) \\ x_1(t) \\ x_2(t) \\ x_0(t) \\ x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_0(t) \\ x_1(t) \\ x_2(t) \\ x_0(t) \\ x_1(t) \\ x_2(t) \end{bmatrix} \tag{1 1 24}$$

which is of the form of equation (1 1 17) Also, if Y_n is the vector of

observations made directly on the position coordinates of the body at time instant t_n, then the observation relation would be

$$
\begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}_n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{0n} \\ x_{1n} \\ x_{2n} \\ x_{0n} \\ x_{1n} \\ x_{2n} \end{bmatrix} + \begin{bmatrix} v_0 \\ v_1 \\ v_2 \end{bmatrix}_n \quad (1\ 1\ 25)
$$

where $Y_n = [y_0 \ y_1 \ y_2]_n^T$ and $[v_0 \ v_1 \ v_2]_n^T$ is the measurement error at the observation instant t_n. Equation (1 1 25) is seen to be of the form of equation (1 1 21)

Alternatively, if Y_n denotes the observations obtained by a radar, located at the Cartesian origin of the range, azimuth and elevation of the body, then the observation relation would be

$$
y_{0n} = (x_{0n}^2 + x_{1n}^2 + x_{2n}^2)^{1/2} + v_{0n} \quad (1\ 1\ 26)
$$

$$
y_{1n} = \tan^{-1}(\frac{x_{1n}}{x_{2n}}) + v_{1n} \quad (1\ 1\ 27)
$$

$$
y_{2n} = \tan^{-1} \frac{x_{2n}}{(x_{0n}^2 + x_{1n}^2)^{1/2}} + v_{2n} \quad (1\ 1\ 28)
$$

which is of the form of equation (1 1 23)

In any modeling process, errors arise from two sources, namely, errors in the representative model itself and errors in the observations The former are known as *systematic errors* and the latter are assumed to be random

A first step in modeling in this framework is to choose a functional representation that adequately describes the process under observation The efficacy of polynomials as functional representations is based on their ability to approximate any continuous function over a finite interval to any degree of

precision

Assume that the process under observation is being modeled by a polynomial x(t) of degree k Denoting the jth- order differentiation by D^j, one can choose a vector

$$X_n = [\ x_n \quad Dx_n \qquad D^k x_n \ ] \tag{1 1 29}$$

Thus, for any n, X_n of equation (1 1 29) provides all the information about the state of the assumed form of the process It is accordingly referred to as a *state-vector* of the chosen model For other forms of functional representation, the state-vector should be chosen so that all the information of the assumed state of the process is contained in it More than one choice of state-vectors is possible For example, for the above one can also choose delayed samples

$$X_n = [\ x_n \quad x_{n-1} \qquad x_{n-k} \ ] \tag{1 1 30}$$

One can show that the time evolution of the state-vector X_n can be related to previous values by

$$X_{n+m} = \phi_m \ X_n \tag{1 1 31}$$

where ϕ_m is known as the *transition matrix* whose specific form depends on the functional representation chosen

In dynamical systems terminology, ϕ_m is called the *flow* of the system in a phase space of dimension k

In equation (1 1 31), ϕ_m provides a linear transformation However, in the case of a non-linear differential equation given by (1 1 19), the flow is also a non-linear transformation and therefore, a linear representation of the type given by equation (1 1 31) cannot be written The general approach then is to linearize the differential equation which is based on the assumption that a *nominal trajectory* $\bar{X}(t)$ is available sufficiently close to the original one X(t),

ıe, the difference vector $\delta X(t)$ ıs sufficiently small so that terms ınvolvıng products and squares of the components of that vector can be ıgnored

Specifically, then one has the following three successively more general situations

a Constant coefficient linear model -

$$X_{n+m} = \phi_m X_n \tag{1 1 31}$$

b Tıme-varyıng linear model -

$$X_{n+m} = \phi_{m,n} X_n \tag{1 1 32}$$

c Non-linear model -

$$X_n = \bar{X}_n + \delta X_n$$

$$\delta X_{n+m} = \phi_{m,n,\bar{X}} \ \delta X_n \tag{1 1 33}$$

The process of observation ıntroduces an *observable* Y_n as gıven by equations (1 1 21)-(1 1 23)

The problem of modelıng then ıs to fınd an approximatıon $\tilde{\phi}$ to the flow ϕ from the observables Y_n The crıterıa usually employed to fınd the approximant $\tilde{\phi}$ ıs to mınımıze the squared error between the estımated state-vector $\tilde{X}_n$ and the actual state-vector X_n, $1 \le n \le N$, ıe, mınımıze

$$E_N = \sum_{n=1}^{N} | X_n - \tilde{X}_n |^2 \tag{1 1 34}$$

where

$$\tilde{X}_{n+1} = \tilde{\phi} (X_n) \tag{1 1 35}$$

and X_n ıs reconstructed from the observables Y_n

State-space modelıng schemes are also very popular ın dıverse fıelds eg ın controls and communıcatıons area ın electrıcal engıneerıng Determınıstıc state-space modelıng schemes can also be used to get an idea of the physıcal laws governıng the phenomenon beıng modeled

The viewpoint that complexity may often arise out of low dimensional chaos leads to new approaches in signal modeling which have been pursued in the thesis The following section discusses some of the aspects informally

1.2 Signal Modeling in the Paradigm of Chaos

As seen in Section 1 1, most of the modeling schemes are linear Even when the functional representation is non-linear, the attempt is to linearize using nominal trajectories etc This naturally breeds the misconception that complex behaviour can arise only out of complex systems, which is true when the additional constraint of linearity is employed Until recently, the only tool to analyse complex behaviour (eg a time series having a nearly flat spectrum etc) was based on Kolmogorov's theory of random processes However, now it is known that simple non-linear systems are capable of extremely complex behaviour A new tool to analyse complex behaviour deterministically is chaos While most of the earlier studies in chaos were directed towards finding regions of chaotic behaviour in specific non-linear systems, some of the recent studies have focussed on the ability to model and predict complex behaviour using the theoretical background developed for analysing chaotic behaviour

Dissipative dynamical systems often have the property that undisturbed trajectories (or the evolution of the flow in state-space) approach a subset of the state-space called an attractor This causes a drastic reduction in the number of degrees of freedom Fluid flows, for example, have an effectively infinite dimensional state-space but can have low dimensional attractors [28]

Thus, it is not important to distinguish chaos from randomness but rather systems with low dimensional attractors from those with high dimensional attractors If a time series is produced by motion on a very high dimensional

attractor, then from a practical point of view it is impossible to gather enough information to exploit the underlying determinism On the other hand, if the attractor dimension is low, then it means that the motion has but a few degrees of freedom and therefore, can be modeled using few independent variables

The important first step in modeling in this framework then is to find the dimension of the attractor In most situations even complex systems have low dimensional attractors, but if the attractors are high dimensional then the probabilistic approach may be as good as any - may be even better and linear models may be optimal The next step is to reconstruct an appropriate state-vector from the observations A powerful tool to aid this was proposed by Packard et al, [59] and put on firm mathematical foundation by Takens [60] Informally put, it states that it is necessary to only observe one variable of the system evolving with time The state-vector may be reconstructed from this variable and the reconstructed state-space will have the same invariant properties eg attractor dimension etc as the state-space of the actual system from which the single variable was observed Thereafter, one may choose any non-linear functional representation (at the present juncture, the choice is adhoc in the absence of greater understanding) for estimating the flow in reconstructed space The choice of a non-linear function is important to accomodate the ability to model apparently random behaviour as well which in this viewpoint is due to chaos rather than intractable complexity Then simple non-linear techniques may be employed to get the parameters of the model

1.3 Scope and Organization of the Thesis

The above ideas of signal modeling using deterministic chaos have been applied to the specific case of speech signals The vocal tract is the underlying

dynamical system in this case We analysed speech signals in terms of phoneme utterances because all meaningful sounds can be constructed from a combination of phonemes Dimensional analysis showed that phonemes are indeed generated by low dimensional attractors Also, the positive values of the second-order entropy showed that most phoneme time series are chaotic in nature Thereafter, we considered two modeling techniques, namely, the *global approximation technique* and the *compromised overlapping neighbourhood (CON) local approximation technique* using polynomials as functional representations and the above ideas We compared the prediction properties of these two techniques with the LPC (covariance-method) which is extensively used in speech processing The global approximation and the CON - local approximation techniques have successively better prediction properties than the LPC

The thesis has been organized as follows

Chapter 2 begins with a review of dynamical systems and their steady-state behaviour including chaos There are various descriptions of the above topics, but a topological description is adopted because it is felt that it provides the best insight to the signal modeling problem The appropriate definitions of the topological terms are introduced as and when they occur This chapter also discusses the various dimensions and entropies that are used to characterize attractors and 'strange' or chaotic behaviour Finally, Takens' theorems are presented which are central to the technique of reconstructing the state-space from a single observed variable of the phenomenon

Chapter 3 begins with a description of the vocal tract as a dynamical system followed by a brief discussion on the classification of phonemes Next, some practical aspects for obtaining the correlation dimension are discussed

followed by their application to phoneme time series Then, the phoneme time series are analysed for their second-order entropy The small values of correlation dimension show that the underlying attractors are low-dimensional while the positive values of second-order entropy show the chaotic nature of phonemes

Chapter 4 begins by building up a theoretical framework of the signal modeling problem Various practical aspects are then considered Some choices of functional representations and two approximation techniques namely the global and local approximation techniques are presented Since the local approximation technique is not practical because of a prohibitively large model order, a compromised overlapping neighbourhood local approximation technique is proposed It has lower model order than the usual local approximation technique and yet attempts to retain the better prediction properties of the local approximation technique over the global approximation technique Finally, a comparison of the computational complexity of various schemes is also made

Chapter 5 concludes the thesis with an overview of the work done and suggestions for future work

CHAPTER 2

REVIEW OF THEORETICAL ASPECTS

This chapter begins with a definition of dynamical systems and their classification in terms of steady state solutions and limit sets It discusses attractors in terms of dimensions and entropies and presents the embedding theorems which are used for reconstructing the state-space from a single observed variable This chapter thus forms the basis of the theoretical framework for the signal modeling problem We begin with dynamical systems

2.1 Dynamical Systems

Dynamical systems are often described in terms of the real space In such a case, a dynamical system with a continuous time evolution can be specified by a system of differential equations Three successively more general classes of differential equations were given by equations (1 1 17)-(1 1 19) and discussed therein Dynamical systems may be grouped into autonomous and non-autonomous systems

A kth- order *autonomous dynamical system* is defined by the state equation

$$ X = f(X), \qquad X(t_0) = X_0 \tag{2 1 1} $$

where $X \in \mathbf{R}^k$ is the state at time t and is a point in the phase space $f \quad \mathbf{R}^k \rightarrow \mathbf{R}^k$ is called the *vector field* which associates a *tangent vector* with each point in phase space The solution to equation (2 1 1) with initial condition X_0 at time t_0 is called the *trajectory* which lies in the phase space and is denoted by $\phi_t(X_0)$ The mapping $\phi_t \quad \mathbf{R}^k \rightarrow \mathbf{R}^k$ is called the *flow* of the system The dynamical system given by equation (2 1 1) is linear if f(X) is linear

A kth- order *non-autonomous dynamical system* is defined by the time-varying state equation

$$X = f(X, t), \quad X(t_0) = X_0 \tag{2.1.2}$$

The vector-field f depends on time. The solution to equation (2.1.2) passing through X_0 at time t_0 is $\phi_t(X_0, t_0)$. The system is linear if f is linear with respect to X.

If there exists a T > 0, such that $f(X, t) = f(X, t+T)$ for all X and all t, the system is said to be *time-periodic* with period T. The smallest such T is called the *minimal period*. An nth- order time periodic non-autonomous system can always be converted to an $(n+1)^{th}$- order autonomous system by appending an extra state $\theta = 2\pi t/T$. The autonomous system will then be given by

$$X = f(X, \theta T/2\pi), \quad X(0) = X_0$$

$$\theta = 2\pi / T \qquad \theta(0) = 2\pi t_0/T \tag{2.1.3}$$

Discrete-time dynamical systems can also be described along similar lines. Any map $f\ \mathbf{R}^k \rightarrow \mathbf{R}^k$ defines a discrete-time dynamical system by the state equation

$$X_{k+1} = f(X_k), \qquad k = 0, 1, \tag{2.1.4}$$

where $X_k \in \mathbf{R}^k$ is called the *state* and f maps the state X_k to the next state X_{k+1}.

Most of the development pertaining to signal modeling is rooted in topological considerations. Consequently, the presentation of dynamical systems in terms of arbitrary manifolds is desirable. A description of some of the terms used in the presentation is now given, although a minimum knowledge is also assumed.

A function is said to be *class k* or C^k if its first k partial derivatives exist and are continuous. If, for a function f $X \rightarrow Y$, there is only one x such

that $y = f(x)$, then the map is said to be *one-to-one* or *injective* If the range of f is the entire space Y, then the map is said to be *onto* or *surjective* and is written as $f(X) = Y$ If f is both one-to-one and onto, then it is *bijective* A *diffeomorphism* is a C^k- bijection A *homeomorphism* is a bijective map f which is bicontinuous (i e , f and f^{-1} are continuous)

An m-dimensional surface in $\mathbf{R}^k$ is defined by relations between the coordinates of $\mathbf{R}^k$ However, this notion can be generalized to a manifold which does not need to be placed in $\mathbf{R}^k$ A *manifold* is a metric space for which every point has a neighbourhood homeomorphic to $\mathbf{R}^k$ This homeomorphism provides coordinates for the manifold by identifying elements of the manifold with the corresponding elements in $\mathbf{R}^k$ The *topological dimension* of a manifold M is that of the corresponding real space A *submanifold* is a manifold contained in another one

An *open cover* of a set S is a collection of open subsets of S such that S is contained in their union A *subcover* is a subset of a cover that is itself a cover A *finite cover* is a cover with a finite number of elements The set S is said to be *compact* if every open cover of S has a finite subcover eg bounded, closed subsets of $\mathbf{R}^k$ are compact

In the same way that the derivative of a function provides its best linear approximation, the best linear approximation to a manifold M at a point x is given by its *tangent space*, denoted by $T_x(M)$ For example, consider a 2-d manifold M placed in $\mathbf{R}^3$ In this case, the tangent space is just a plane Let C be a curve in M (C $\mathbf{R} \rightarrow \mathbf{R}^3$, C $t \rightarrow x$) The *tangent vector* to C at x is a vector in $\mathbf{R}^3$ given by

$$v_x C = \frac{dx}{dt} \tag{2 1 5}$$

The set of tangent vectors to all curves passing through a point is the tangent space to the manifold at that point

Following is a description of dynamical systems in terms of arbitrary manifolds

Let the phase space of a system be a compact manifold M A dynamical system on M is a map ϕ M $\rightarrow$ M (discrete-time) or a vector-field (continuous-time) The vector-field associates each point in the manifold with an element of the tangent space at that point The time evolution of a dynamical system is given by the *flow* ϕ^1 (X_0) (discrete-time) or ϕ_t (X_0) (continuous-time) In experimental situations, the process of sampling the continuous-time flow ϕ_t introduces the discrete-time map ϕ^1 i e, $\phi^1 = \phi_{1\tau}$, τ being a fixed delay and i $\in$ N An *observable* on the dynamical system is a smooth funtion y M $\rightarrow$ R

2.2 Steady-State Behaviour, Limit Sets and Attractors

The steady-state behaviour of a dynamical system refers to its asymptotic behaviour Only bounded steady-state responses are of interest to us Some definitions are given before a discussion of the steady-state behaviour of dynamical systems

An *invariant set* S of a flow ϕ on a manifold M is a subset of M defined by

$$S = \{ X \in M / \phi_t (X) \in S \ \forall \ X \in \ S \ and \ \forall t \} \qquad (2\,2\,1)$$

A point Y is a *limit point* of X if for every neighbourhood U of y, ϕ_t (X) repeatedly enters U as t $\rightarrow$ ∞

The *positive limit sets* (or *limit cycle*) L^+ (X) of a point X is the set of limit points that the flow approaches with an initial condition of X

$$L^+ (X) = \{ Y \in M / \exists \ t_1 \rightarrow \infty \ with \ \phi_{t_1} (X) \rightarrow Y \} \qquad (2\,2\,2)$$

Negative limit sets are defined for flows going backward in time Limit sets are closed and invariant under ϕ_t

A limit set L is an *attracting set* if it has a neighbourhood U such that $\phi_t(X) \in U \quad \forall$ t and $\phi_t(X) \rightarrow S$ (i e, the flow comes arbitrarily close to all members of S) as $t \rightarrow \infty$

The *basin of attraction* B (L) of an attracting limit set L is defined as the union of all such neighbourhoods U Every trajectory in B (L) tends towards L as t $\rightarrow \infty$

Limit sets are useful in describing the classical types of steady-state behaviour eg equilibrium points, limit cycles, quasi-periodic attractor etc but the definition cannot be extended to the complex steady-state behaviour found in chaotic systems The term *strange attractor* is used to describe the object on which the trajectories of a chaotic system accumulate Four different types of steady-state behaviour follow

1 Equilibrium points

- An equilibrium point X_{eq} of an autonomous system given by equation (2 1 1) is a constant solution

$$\phi_t (X_{eq}) = X_{eq} , \text{ for all } t \qquad (2\,2\,3)$$

At the equilibrium point the vector field or map vanishes i e, f $(X_{eq}) = 0$ An example is the damped pendulum equation

$$X = Y$$
$$Y = -kY - \sin X \qquad (2\,2\,4)$$

which has an infinity of equilibrium points at $(X,Y) = (k\pi, 0)$, k= 0, ± 1, ± 2, which constitutes the invariant set The limit set for an equilibrium point is the equilibrium point itself

2 Periodic Solutions

:- $\phi_t (X^*, t_0)$ is a periodic solution if

$$\phi_t (X^*, t_0) = \phi_{t+T'} (X^*, t_0) \qquad (2\,2\,5)$$

for all t and some minimal $T' > 0$ A periodic solution has a Fourier transform consisting of a fundamental component at $f = 1/T'$ and evenly spaced harmonics at

k/T', $k = 2,3$,   The amplitude of some of these spectral components may be zero

Figure 2.1 shows the fundamental and subharmonic solutions together with their spectrum for the Duffing's equation

$$X = Y$$
$$Y = X - X^3 - \epsilon Y + \tau \cos \omega t \qquad (2.2.6)$$

In the autonomous case, an isolated periodic solution[1] ϕ_t (X^*) is called a *limit cycle* A limit cycle is a self-sustained oscillation and cannot occur in a linear system An example of a limit cycle is found in the van der Pol equation

$$X = Y$$
$$Y = (1 - X^2)Y - X \qquad (2.2.7)$$

Figure 2.2 shows the trajectory, time waveform and the spectrum of limit cycle behaviour

The limit set corresponding to a limit cycle is the closed curve traced out by ϕ_t (X^*) over one period

3 Quasi-Periodic Solutions -- Such a solution can be described by a set of k incommensurate frequencies ω_1, ω_2,   , ω_k, i e,

$$X (t) = f (\omega_1 t, \omega_2 t,   , \omega_k t) \qquad (2.2.8)$$

The motion equation (2.2.8) describes is a k-dimensional torus T^k (i e., a product of k circles) embedded in $\mathbf{R}^k$ and constitutes a quasi-periodic attractor Quasi-periodic waveforms may be created when two or more periodic functions interact non-linearly

Quasi-periodic solutions may be described with the example of the van der Pol equation (2.2.7) It possesses a limit cycle whose natural period T_1 depends on the system parameters Now add a sinusoidal term as follows

---

[1] A periodic solution is *isolated* if there exists a neighbourhood of it that contains no other periodic solution
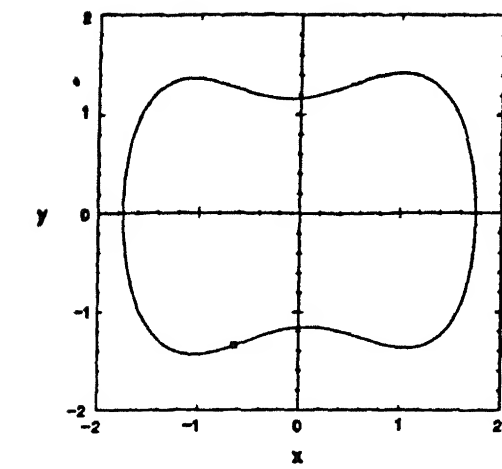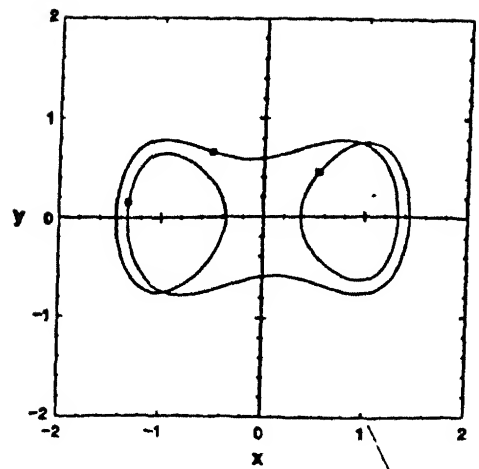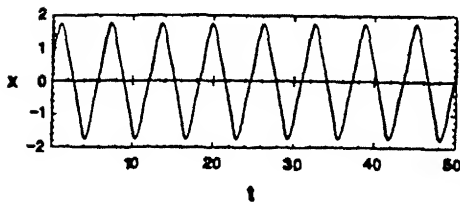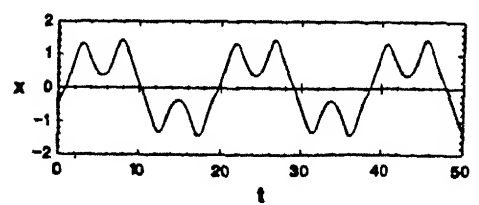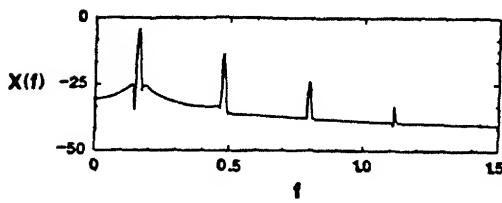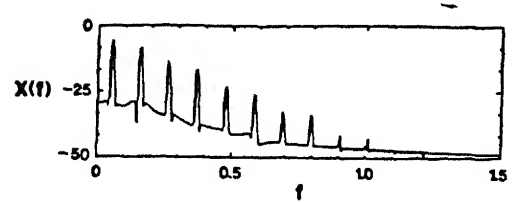
Figure 2.1 - Periodic solutions of Duffing's equation for $\tau = 0\,3$, $\omega = 1$. a) Period 1 solution. b) Time waveform of a), c) Spectrum from b), d) Period 3 subharmonic, e) Time waveform of d), f) Spectrum from e). Only odd harmonics are present in both c) and f). (Reproduced from [20])
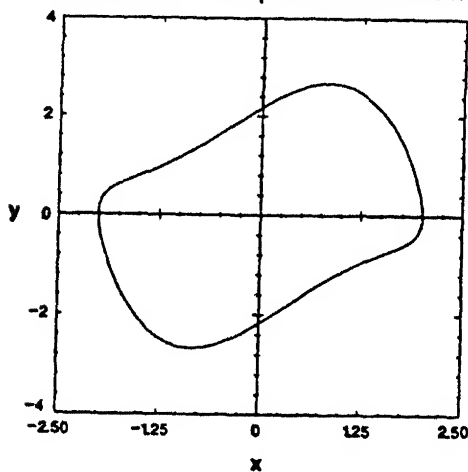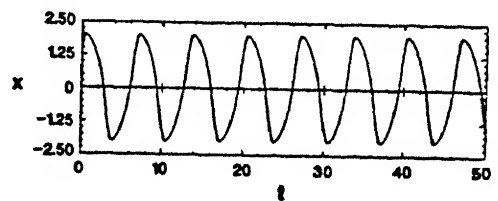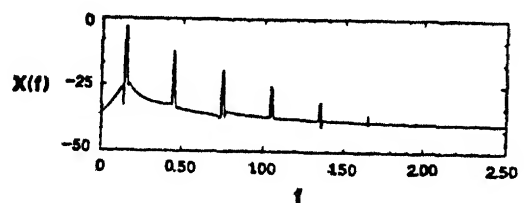
Figure 2.2 - Limit cycle for the van der Pol equation a) Trajectory, b) The time waveform, c) Spectrum of b) - due to symmetry of the time-waveform (Reproduced from [20])

$$X = Y$$

$$Y = (1 - X^2)Y - X + A \cos (2\pi t/T_2) \tag{2 2 9}$$

The solution to the forced system could synchronize with some multiple of the input period T_2 resulting in a subharmonic Alternatively, in a conflict between T_1 and T_2, neither period wins, and quasi-periodic behaviour results A two-periodic trajectory of the forced van der Pol equation (2 2 9) along with its time waveform and spectrum is shown in figure 2 3

4 Chaos and Strange Attractors

- There is no generally accepted definition of chaos From a practical point of view chaos can be described as bounded steady-state behaviour that is not an equilibrium point, not periodic and not quasi-periodic In the following, an attempt to describe chaotic behaviour is made and then a definition based on topological considerations is given

An example of a chaotic trajectory along with its time waveform and spectrum is shown in figure 2 4 It is evident from the figure that the trajectory is bounded, but not periodic and does not have the uniform distribution characteristic of quasi-periodic solutions The limit set for chaotic behaviour is not a set of simple geometrical objects like a circle or torus, but is related to fractals and Cantor sets [19] Another property of chaotic systems is sensitive dependence on initial conditions i e, given two different initial conditions arbitrarily close to one another, the trajectories emanating from these points diverge at a rate characteristic of the system until for all practical purposes, they become uncorrelated (see figure 2 5) In practice, the initial state of the system can never be specified exactly, but only to within some tolerance $\xi > 0$ If two initial conditions X_0 and $\hat{X}_0$ lie within ξ of one another, they cannot be distinguished However, after a finite amount of time, $\phi_t (X_0)$ and $\phi_t (\hat{X}_0)$ will diverge and become uncorrelated Therefore, no matter how precisely the initial
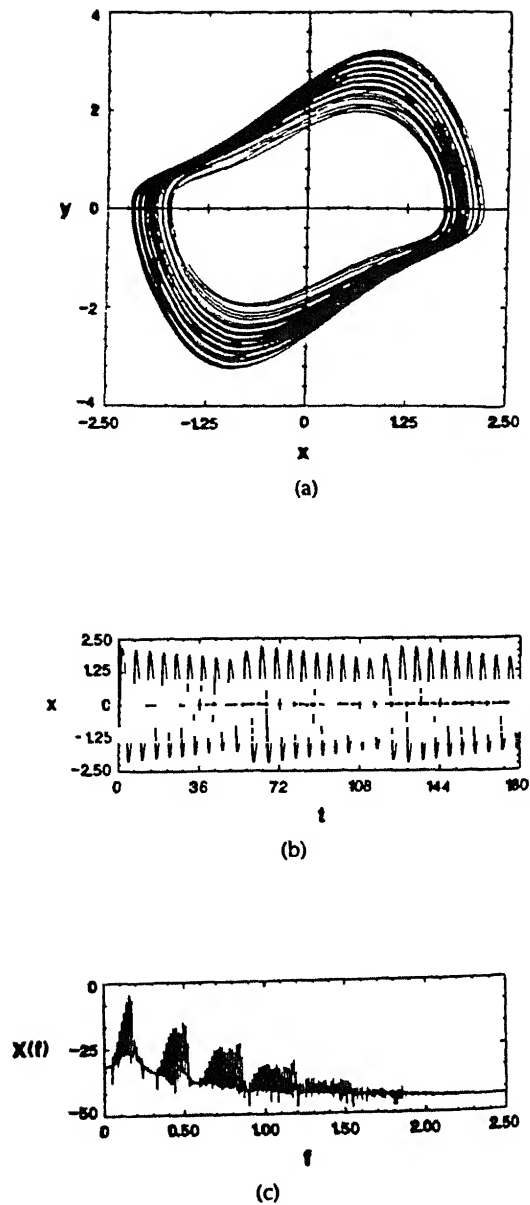
Figure 2 3 - Quasi-periodic behaviour of the non-autonomous van der Pol equation for A=0 5 and $T_2=2\pi/1\ 1$. a) The trajectory, b) The time waveform, c) Spectrum of b). (Reproduced from [20])

Figure 2.4 - a) Chaotic trajectory of a second-order non-autonomous system, b) Time waveform, c) Spectrum of b) (Reproduced from [20])
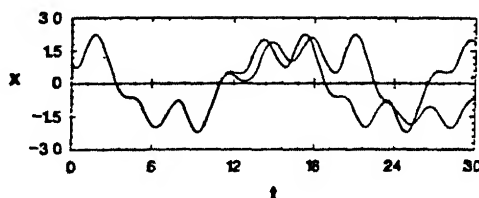
Figure 2 5 - Two trajectories plotted versus time showing sensitive dependence on initial conditions in a third order autonomous system The initial conditions differ by 0.01 %. (Reproduced from [20])

condition is known, the long term behaviour of chaotic systems cannot be predicted That is what is meant by the apparently random behaviour of such deterministic systems Many such systems are exploited to generate the so called pseudo-random numbers

Again, some terms are presented before giving a definition of chaos based on the topological approach [16] The discussion is terms of a discrete-time system, although it easily extends to the continuous-time case by analogy

For any set S, its *closure* $\bar{S}$ consists of all points in S together with all limit points of S (For definition of limit point, see earlier part of this section) A subset U of S is *dense* in S if $\bar{U} = S$

A function f J $\rightarrow$ J is said to be *topologically transitive* if for any pair of open sets U,V $\subset$ J, $\exists$ k $>$ 0 such that f^k (U) $\cap$ V $\neq$ 0 A map possesing a dense orbit (equivalent to a trajectory in a flow) is topologically transitive Intuitively, a topologically transitive map has points which eventually move under iteration from one arbitrarily small neighbourhood to any other

A function f J $\rightarrow$ J has *sensitive dependence on initial conditions* if $\exists$ ε $>$ 0 such that for any x and for any neighbourhood N of x, $\exists$ y $\in$ N and n $>$ 0 such that $| f^n(x) - f^n(y) | > \varepsilon$

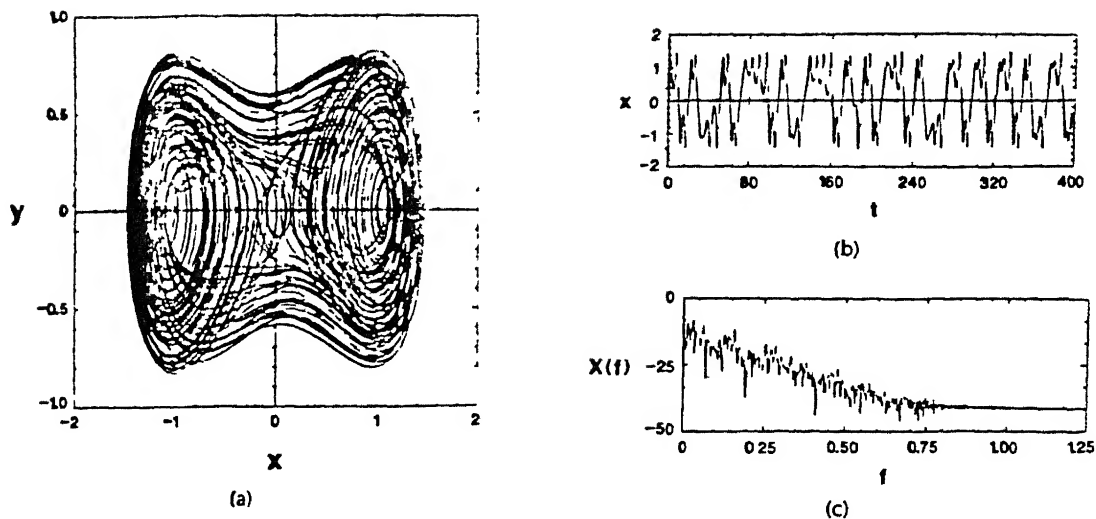A map possesses sensitive dependence on initial conditions if there exist points arbitrarily close to x which eventually separate from x by at least ε under iteration f It is important to note that not all points near x need eventually separate but at least one such point in every neighbourhood should

A chaotic map can now be defined as follows

Let V be a set f V $\rightarrow$ V is said to be *chaotic* on V if

i   f has sensitive dependence on initial conditions,

ii   f is topologically transitive,

iii   the periodic points are dense in V

Thus, chaotic maps are indecomposible and they possess an element of regularity amidst long term unpredictability The unpredictability is due to sensitive dependence on initial conditions The map cannot be decomposed into two subsystems (i e , two invariant open sets) which do not interact under f because of topological transitivity The element of regularity is because of periodic points which are dense

An operational definition of a strange attractor can now be given A *strange attractor* is an attracting set possessing properties of invariance, topological transitivity and sensitive dependence on initial conditions It is the sensitive dependence on initial conditions that makes the attractor strange

Lyapunov exponents are a convenient way of categorizing steady-state behaviour[49,51,54] Given a continuous dynamical system in a k-dimensional phase space, the long term evolution of an infintesimal k-sphere of initial conditions is monitored The sphere becomes a k-ellipsoid due to the locally deforming nature of the flow The ith 1-dimensional Lyapunov exponent is defined in terms of the length of the ellipsoidal principal axis $p_i(t)$

$$\lambda_i = \lim_{t \to \infty} \frac{1}{t} \log_2 \frac{p_i(t)}{p_i(0)}$$

(2 2 10)

Thus, Lyapunov exponents are related to the expanding or contracting nature of different directions in phase space For an attractor, contraction must outweigh expansion, so that

$$\sum_{i=0}^{n} \lambda_i \; < \; 0$$

The Lyapunov exponents are generally written in decreasing order i e , $\lambda_1 \geq \lambda_2 \geq \lambda_3$

For a stable equilibrium point, $\lambda_i < 0$ for all i For a stable limit cycle, $\lambda_1 = 0$ and $\lambda_i < 0$ for $i = 2,3,$, n For a stable quasi-periodic attractor with k rationally independent frequencies, $\lambda_1 = 0,$, $\lambda_k = 0, \lambda_{k+1} < 0,$ A strange attractor has at least one positive Lyapunov exponent Extraction of Lyapunov exponents from experimental time series is difficult although partial attempts in this direction have been made [50,53,55]

2.3 Attractor Dimensions and Entropies

In this section, we study the classification of attractors and trajectories using the concepts of dimensions and entropies [18,20,31]

An attractor is said to be k-dimensional if in a neighbourhood of every point, it is diffeomorphic to an open subset of $\mathbf{R}^k$ For example, a limit cycle is one-dimensional since it looks locally like an interval A torus is two-dimensional since it locally resembles an open subset of $\mathbf{R}^2$ An equilibrium point has zero dimension The neighbourhood of any point of a strange attractor has a fine structure which does not resemble any Euclidean space They have non-integer dimension

There are various generalized dimensions to cope with non-integral situations eg capacity and Hausdorff dimension, information dimension, correlation dimension etc Instead of presenting them individually, a unified procedure is adopted [38,45]

The discrete-probability is given by $p_i = N_i/N$, where N is the total number of elements in the considered sample space Given some partition Φ consisting of

elements $(\Phi_1, \quad , \Phi_m)$ in the sample space, one can count the number of times N_1 an element is found in Φ_1 The qth- order information H_q is defined as

$$H_q = \frac{1}{1-q} \log \sum_1 p_1^q \qquad (2\ 3\ 1)$$

Now consider the basin of attraction as a sample space and take a finite partition Φ (r) with diameter r of this basin in which the trajectory X(t) is situated Then,

$$H_q (r) = \inf_{\Phi(r)} H_q (\Phi(r)) \qquad (2\ 3\ 2)$$

denotes a qth- order information which depends on the partition diameter r $H_q(r)$ is given by the infimum of the different informations resulting from all possible partitions Φ (r)

One can define a quantity D_q of order q given by

$$D_q = - \lim_{r \to 0} \frac{\log H_q (r)}{\log r} \qquad (2\ 3\ 3)$$

which is the generalized q^{th}- order dimension of the attractor

The *fractal dimension* or *capacity* of the attractor is $D_0 = \lim_{r \to 0} D_q$ and given by

$$D_0 = - \lim_{r \to 0} \frac{\log M (r)}{\log r} \qquad (2\ 3\ 4)$$

It is determined by M (r), the minimal number of cubes with edge length r needed to cover the attractor Equation (2 3 4) is equivalent to Mandelbrot's definition of the fractal dimension [19] which originates from Hausdorff For manifolds, D_0 is equal to the dimension of the manifold which is an integer However, for objects with fractal structure, D_0 is usually a non-integer The following two examples illustrate this

Example 1 - The unit interval   Cover the unit interval [0,1] with volume elements (intervals) of length $r = 1/3^n$ (Figure 2.6(a)) Then, $M(r) = 3^n$ To refine the covering, let $n \rightarrow \infty$ to arrive at

$$D_0 = -\lim_{n \to \infty} \frac{\ln 3^n}{\ln 1/3^n} = 1 \qquad (2.3.5)$$

Thus, the unit interval has dimension 1 as expected

Example 2 - The middle-third Cantor set   Remove the middle third of the unit interval leaving the two intervals [0,1/3] and [2/3,1] Remove the middle third of each of these intervals leaving four intervals (see figure 2.6(b)) Repeat this process ad infinitum The resulting set is called the *middle-third Cantor set* They have interesting properties like fractional dimension and a fine structure similar
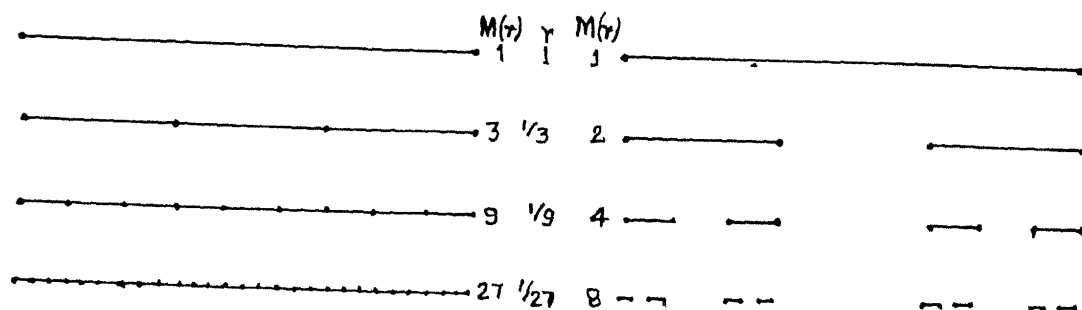
Figure 2.6 - Two examples of the fractal dimension a) The unit-interval, b) The middle-third Cantor set.

to that observed in chaotic systems To show that D_0 of this Cantor set is not an integer, choose a covering of intervals with length $r = 1/3^n$ Then, $M(r) = 2^n$ and,

$$D_0 = \lim_{n \to \infty} \frac{\ln 2^n}{\ln 3^n} = \frac{\ln 2}{\ln 3} = 0.6309 \qquad (2.3.6)$$

Hence, the Cantor set is something more than a point (dimension = 0), but something less than an interval (dimension = 1)

The *information dimension* [32] of the attractor is $D_1 = \lim_{q \to 1} D_q$, and given by

$$D_1 = -\lim_{r \to 0} \frac{S(r)}{\log r} \qquad (2.3.7)$$

where

$$S(r) = -\sum_{i=1}^{M(r)} p_i \log p_i \qquad (2.3.8)$$

is the *entropy* i e, the amount of information needed to specify the state of the system to an accuracy r if the state is known to be on the attractor Hence, the name information dimension While D_0 does not utilize any information about the time behaviour of dynamical systems, D_1 takes into account the relative frequency of visitation of a trajectory in a cube of edge length r

The *correlation dimension* [35] of the attractor is $D_2 = \lim_{q \to 2} D_q$, and given by

$$D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log r} \qquad (2.3.9)$$

where

$$C(r) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=1}^{N} \theta(r - |X_i - X_j|) \qquad (2.3.10)$$

θ being the Heaviside function $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ otherwise It has been shown that $D_2 \leq D_1 \leq D_0$ The conditions for equality of the dimensions are satisfied if the points are distributed uniformly over the attractor D_2 is the easiest to compute from an observed time series[35] and is generally the one that

is computed to get an idea of the attractor dimension

We now consider the definition of qth- order entropy of a trajectory X (t) situated in the basin of the attractor The considered points X (t) along the trajectory are separated by a constant time increment τ The whole phase space is partitioned into cubes of edge length r $p (i_1, i_2, \quad ,i_d)$ is the joint probability that X (t=τ) is in cube i_1, X (t=2τ) is in cube i_2, and X (t=dτ) is in cube i_d The qth- order entropy is defined as

$$K_q = - \lim_{r \to 0} \lim_{d \to \infty} \frac{1}{d\tau(q-1)} \log \sum_{i_1, \, ,i_d} p^q (i_1, \, ,i_d) \qquad (2\ 3\ 11)$$

K_0 is the *topological entropy*

The first-order entropy, $K_1 = \lim_{q \to 1} K_q$ is the *metric or Kolmogorov entropy* which is a measure for the internal information production during its temporal evolution

$$K_1 = \lim_{r \to 0} \lim_{d \to \infty} \frac{1}{d.\tau} \sum_{i_1, \, ,i_d} p (i_1, \, ,i_d) \log p (i_1, \, ,i_d) \qquad (2\ 3\ 12)$$

According to the Pesin identity, K_1 is approximately equal to the sum of the positive Lyapunov exponents of the system

Similar to the second-order dimension D_2, a second-order entropy K_2 can be defined by the $C_d(r)$, the correlation function C (r) for reconstructed dimension d

$$K_2 = \lim_{r \to 0} \lim_{d \to \infty} \frac{1}{\tau} \log \frac{C_d (r)}{C_{d+1} (r)} \qquad (2\ 3\ 13)$$

Some practical points regarding the evaluation of correlation dimension and the second-order entropy will be discussed in Chapter 3 in the context of application to phoneme time series

As in the case of dimensions, it has been shown that $K_2 \leq K_1 \leq K_0$ The limiting cases, $K_1 = 0$ and $K_1 \to \infty$ characterize situations of regular (eg

periodic) and random behaviour respectively $K_1 > 0$ signifies chaotic behaviour

Since $K_2 \leq K_1$, $K_2 > 0$ is a sufficient condition for deterministic chaos

2.4 Reconstruction of Attractors

A system's dynamics can be reconstructed from a single degree of freedom as suggested by Packard et al[59] and put on firm mathematical foundation by Takens[60]

Consider a dynamical system as described in terms of arbitrary manifolds in Section 2 1 It is briefly restated as follows

The system to be considered is a manifold M in which the system state can be defined Consider a map φ M $\rightarrow$ M (discrete-time) or a vector field X on M (continuous-time) which defines the dynamics, the flow φ^1 (X_0) (discrete-time) or φ_t (X_0) (continuous-time). and an observable y M $\rightarrow$ **R** Takens showed that M can be embedded in $\mathbf{R}^k$ through the observable y, the limit set of the flow is reproduced and that the attractor dimension is unchanged by the embedding

In the following the relevant theorems due to Takens are stated without the proofs

Theorem (Takens) - Let M be a compact manifold of dimension m For pairs (φ,y), φ M $\rightarrow$ M, a smooth diffeomorphism, and y M $\rightarrow$ **R**, a smooth function, it is a generic property that the map $\Phi_{(\varphi,y)}$ M $\rightarrow$ $\mathbf{R}^{2m+1}$, defined by

$$\Phi_{(\varphi,y)} (X) = (\ y(X),\ y(\varphi(X)),\ y(\varphi^2(X)),\quad ,y(\varphi^{2m}(X))\ ) \qquad (2\ 4\ 1)$$

is an embedding, by *smooth* we mean at least C^2

For each point X $\in$ M, Φ (X) gives a point in $\mathbf{R}^{2m+1}$, and this mapping is an embedding While 2m+1 is the largest real space needed to fit the manifold, it often turns out that the dimension can be less For only a pathological or

trivial choice of M, y and φ will the derivatives of the iterates of the map be linearly dependent If such is the case, then an arbitrary perturbation in the space of possible manifolds, functions or observables will remove the linear dependence In practice, the presence of noise guarantees this perturbation

Takens also showed that instead of time delays, 2m+1 derivatives will also work It is an open question whether other choices for the set of functions might in some sense be optimal However, time delays are the most commonly used functions due to their simplicity

Generally, sampled versions of signals are observed in practical situations The process of sampling introduces a discrete-time map φ_τ from the continuous-time flow φ_t Takens has shown that the limit sets for the continuous-time flow and the discrete-time map derived from the flow are identical for most choices of time increments *Most* means a residual (open and dense) subset of time increments

Theorem (Takens)

Let M be a compact manifold, X a vector field on M with flow φ_t and P a point in M Then, there is a residual subset $C_{X,P}$ of positive real numbers such that for $\tau \in C_{X,P}$, the positive limit sets of P for the flow φ_t of X and for the diffeomorphism φ_τ are the same In other words, for $\tau \in C_{X,P}$, we have that each point $q \in M$ which is the limit of the sequnce $\varphi_{t_1}(P)$, $t_1 \in \mathbb{R}$, $t_1 \to \infty$, is the limit of a seqence $\varphi_{n_1\tau}(P)$, $n_1 \in \mathbb{N}$, $n_1 \to \infty$

These two theorems pave the way for a corollary that finally gives the central result on observing dynamical systems In the above discussion and theorems, it was shown that discrete-time maps on the manifolds lead to an embedding in $\mathbb{R}^{2m+1}$ through the observable and time delays, and that the

conversion of a continuous-time flow into a discrete-time map by choosing a fixed sampling interval gives the same limit set These results together imply that the discretely-sampled data set embedded into $\mathbf{R}^{2m+1}$ has the same limit set as the real system

Corollary (Takens)   Let M be a compact manifold of dimension m We consider quadruples, consisting of a vector field X, a function y, a point P and a positive real number τ For generic such (X,y,P,τ) (more precisely for generic (X,y) and τ satisfying generic conditions depending on X and P), the positive limit set $L^+(P)$ is diffeomorphic with the set of limit points of the following sequence in $\mathbf{R}^{2m+1}$

$$S_{X,y,P,\tau} = (\ (y(\varphi_{k\tau}(P)),\ y(\varphi_{(k+1)\tau}(P)),\quad ,\ y(\varphi_{(k+2m)\tau}(P)))\ )_{k=0}^{\infty} \qquad (2\ 4\ 2)$$

Here diffeomorphic means that there is a smooth embedding of M into $\mathbf{R}^{2m+1}$ mapping $L^+(P)$ bijectively to the set of limit points of this sequence

The final theorem states that the capacity of the limit set of the flow or the real system is the same as the capacity of the embedded flow This follows from the previous observation that dimensions are preserved by embeddings

Theorem (Takens) · $d_C (\ L^+(p)\ ) = d_C [\ L^+(\ S_{X,y,P,\tau})\ ]$ \qquad (2 4 3)

In this chapter, a review of the preliminaries for the signal modeling was given Only those aspects of dynamical systems behaviour and their characterization were discussed which will be relevant later on in the thesis Thus, for example, an otherwise important topic of routes to chaos was not discussed In the next chapter, we will be concerned with the vocal tract as dynamical system and will discuss the practical aspects of dimension and entropy analysis in the context of speech signals

CHAPTER 3

ON SPEECH PRODUCTION AND ATTRACTOR DIMENSION AND ENTROPY OF PHONEME TIME SERIES

This chapter begins with a presentation of the vocal tract as a dynamical system, the mechanism of speech production and its traditional analysis Thereafter, the ideas developed in the previous chapter are used to carry out dimension and entropy analysis in terms of unit utterances, namely phonemes The practical aspects of the analysis scheme are also considered The analysis shows that phoneme time series are generated by low dimensional attractors Moreover, the positive values of second-order entropy show the chaotic nature of phoneme time series These two observations have various implications eg synthesis of phonemes by extracting formant frequencies may not be adequate as indeed the synthesised sounds indicate Moreover, analysis of speech signals in terms of linear systems is inadequate and, in general, long term prediction as well as higher level recognition may indeed be impossible We are therefore motivated to model speech signals using non-linear functional representations which are capable of modeling chaotic behaviour as well These modeling schemes and their performance comparison with the LPC will then be discussed in the next chapter

3.1 Vocal Tract as a Dynamical System and Acoustic Phonetics

The speech production mechanism has the vocal tract as the underlying dynamical system Also, all meaningful sounds can be expressed in terms of phonemes This section briefly discusses the vocal tract and phoneme classification according to the vocal tract characteristics during their utterance For a good discussion one may consult [79]

Figure 3 1 shows an X-ray photograph in which the important features of the vocal tract have been highlighted by dotted lines The *vocal tract* begins at the opening between the *vocal chords* or *glottis* It consists of the *pharynx* or the connection from the esophagus to the mouth, and the mouth or the *oral cavity* The vocal tract ends at the lips The total length of the vocal tract in an average male is about 17 cm The cross-sectional area of the vocal tract determined by the positions of the tongue, lips, jaw and velum varies from zero (i e , complete closure) to about 20 cm^2 The *nasal tract*, which begins at the velum and ends at the nostrils is acoustically coupled to the vocal tract when the velum is lowered to produce nasal sounds of speech

Speech sounds can be classified into 3 classes according to the type of excitation *Voiced sounds* are produced by forcing air through the glottis with the tension of the vocal chords adjusted so that they vibrate in relaxation oscillation Examples are /b/ as in *buy*, /m/ as in *me*, /y/ as in *yellow* etc *Unvoiced* or *fricative sounds* are generated by forming a constriction at some point in the vocal tract, and forcing air through the constriction at some high enough velocity to produce turbulence eg /z/ as in *zip*, /v/ as in *vine* etc *Plosive sounds* result from making a complete closure (usually towards the front end of the vocal tract), building up pressure behind the closure, and abruptly releasing it eg /tʃ/ as in *chip*, /t/ as in *tie* etc

The following discussion deals with the classification of phonemes into four broad classes namely, vowels, dipthongs, semi-vowels and consonants Also, each of these phonemes can be classified as either a *continuant* or *non-continuant* sound Continuant sounds are produced by a fixed vocal tract configuration excited by an appropriate source They include vowels, fricatives (voiced and unvoiced) and nasals The remaining sounds (dipthongs, semi-vowels, stops and
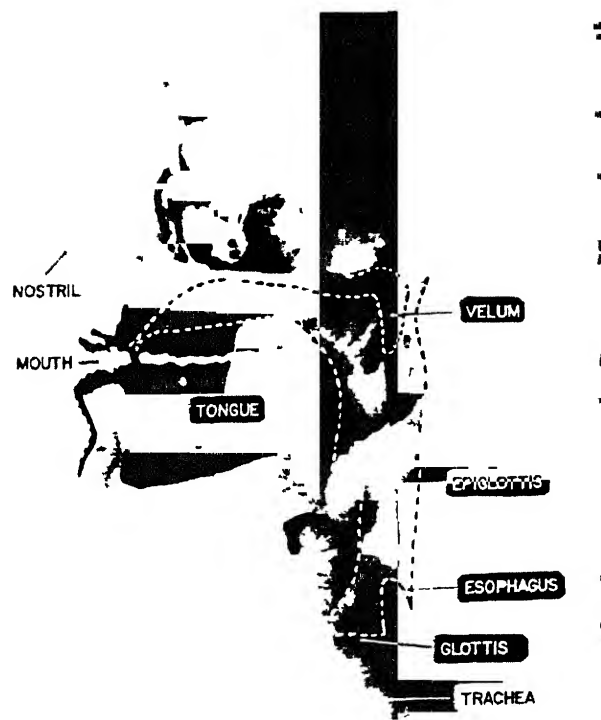
Figure 3 1   An X-ray of the vocal tract system [Reproduced from 79]

affricates) are produced by changing vocal tract configuration and are called non-continuants Figure 2-10, parts a) and b) in each show a part of the time-series and the corresponding spectrum of phonemes /u/, /I/, /r/, /m/, /f/, /v/, /3/, /k/ and /t∫/ which are representative of the following classes

1 Vowels

They are produced by exciting a fixed vocal tract with quasi-periodic pulses of air caused by vibration of the vocal chords The dependence of the cross-sectional area upon the distance along the tract is called the *area function* of the vocal tract For a particular vowel, the area function depends primarily on the position of the tongue, but the positions of the jaw, lips and velum also influence the sound Thus, each vowel sound can be characterized by the vocal tract configuration (or the area function) that is used in its production An alternate approach is to assume the vocal tract as a resonance tube and characterize vowels in terms of the first three resonance frequencies which are known as *formants* in speech terminology Examples of vowels are /i/ as in bead, /I/ as in bid, /ae/ as in bad etc

2 Dipthongs

A dipthong is described as a gliding monosyllabic speech item that starts at or near the articulatory position for one vowel and moves to or towards the position of another The three dipthongs of the International Phonetic Alphabet (IPA) are /aI/ as in ride, /aU/ as in *out* and /ɔI/ as in boy Thus, the dipthongs are produced by varying the vocal tract between appropriate vowel configurations and hence they can be characterized by a time varying area function

3 Semi-Vowels

Similar to dipthongs, these are characterized by a gliding transition in the area function between adjacent phonemes and are strongly influenced by the context in which they occur Examples are /w/ as in *we*, /l/ as in lap and /r/ as in rap

4 Nasals

The nasal consonants /m/ as in *me*, /n/ as in *not* and / / as in th*ing*

are produced with glottal excitation and the vocal tract constricted at some point along the oral passageway The velum is lowered so that air flows through the nasal tract and sound is radiated at the nostrils Nasal consonants and nasalized vowels (i e , some vowels preceding or following nasal consonants) are characterized by resonances which are spectrally broader or more highly damped than those for vowels The broadening of the nasal resonances is due to the fact that the inner surface of the nasal tract is convoluted so that the nasal cavity has a relatively large ratio of surface to cross-sectional area This also means that heat conduction and viscous losses are larger than normal

5 Unvoiced Fricatives

The unvoiced fricatives /f/ as in fun, /s/ as in sap, /θ/ as in thought and /ʃ/ as in ship are produced by exciting the vocal tract by a steady air flow which becomes turbulent in the region of a constriction in the vocal tract

6. Voiced Fricatives

The voiced fricatives /v/ as in vine, /z/ as in zip, /ծ/ as in them and / / as in measure are the counterparts of the corresponding unvoiced fricatives /f/, /s/, /θ/ and / /, in the sense that the place of constriction are the same for both However, voiced fricatives are different from the unvoiced counterparts in that the former have two excitation sources in their production as opposed to one of the latter For voiced fricatives the vocal chords also vibrate and thus the glottis acts as an excitation source

7 Voiced Stops

The voiced stop consonants are /b/ as in bug, /d/ as in dog, /g/ as in got They are transient, noncontinuant sounds and are produced by building up pressure behind a total constriction somewhere in the oral tract and suddenly releasing the pressure These sounds are highly influenced by the vowels which follow the stop consonant By themselves, the waveforms for stop consonants give little information about the particular stop consonant

8 Unvoiced Stops

The voiced stop consonants /p/ as in pie, /t/ as tie and /k/ as in key are similar to their voiced counterparts /b/, /d/ and /g/ One major

difference is that in this case, during the period of total closure of the vocal tract, as the pressure builds up, the vocal chords do not vibrate Thus, following the period of closure, as the air pressure is released, there is a brief interval of friction (due to sudden turbulence of escaping air) followed by a period of aspiration (steady airflow from the glottis exciting the resonances of the vocal tract) before voiced excitation begins

9 Affricates and /h/ The unvoiced affricate /tʃ/ can be modeled as a concatenation of the stop /t/ and the fricative /ʃ/ The voiced affricate /ʒ/ can be modeled as a concatenation of the stop /d/ and the fricative /ʒ/ Finally, the phoneme /h/ is produced by exciting the vocal tract by a steady air flow - i e, without the vocal chords vibrating, but with the turbulent flow being produced at the glottis

This completes the brief discussion on the classification of phonemes The analysis and synthesis of phonemes assume importance in speech processing because all meaningful sounds can be broken down into these unit utterances

3.2 Some Aspects of the Acoustic Theory of Speech Production

In most applications, sound is associated with vibrations of particles in the media and the frequencies of these vibrations Thus, the laws of physics form the basis for describing the generation and propogation of sound In particular, for describing the vocal system, the fundamental laws of the conservation of energy, conservation of momentum along with the laws of thermodynamics and fluid mechanics must be applied to the compressible, low viscosity fluid (air) that is the medium for sound propogation in speech A complete acoustic theory must include the effects of time variation of the vocal tract shape, losses due to heat conduction and viscous friction at the vocal tract

walls, softness of the vocal tract walls, radiation of sound at the lips, nasal coupling and excitation of sound in the vocal tract

A complete theory incorporating all the above effects is not yet available The general approach in building an acoustic theory is to make various simplifying assumptions One of the simplest descriptions is to model the vocal tract as a tube of non-uniform, time-varying cross-section For frequencies corresponding to wavelengths that are long compared to the dimension of the vocal tract, plane wave propogation along the axis of the vocal tract tube is assumed A further simplifying assumption is that there are no losses due to viscosity or thermal conduction either in the bulk of the fluid or at the walls of the tube Such a model is due to Portnoff [78]

For more complicated behaviour, models were proposed [74,80] to account for viscous friction between the air and the walls of the tube, heat conduction through the walls of the tube and vibration of the tube walls A new set of the equations of motion from the first principles is extremely difficult because of the frequency dependence of the losses Hence, the approach is modify the frequency domain representation of the equations of motion to account for the above effects The effects of wall vibration is more pronounced than those of viscous friction and thermal conduction However, viscous and thermal losses increase with frequency and have the greatest effect in the high frequency resonances

Another effect is due to radiation at the lips The models referred to above assume the pressure at the lips to be zero at all times In reality, however, the vocal tract tube terminates with an opening between the lips Hence, a reasonable model should consider the lip as an orifice in a sphere which then acts as a radiating surface with the radiated sound being diffracted by the

spherical baffle that represents the head Again, the resulting diffraction effects are complicated and difficult to represent

It is thus seen that most of the observed effects are difficult to incorporate in physical models derived from first principles In such a situation, one resorts to phenomenological models where the observed behaviour is approximately modeled using linear functional representations that are not related to the physical laws governing the system In such a representation the system output has the desired speech like properties when controlled by a set of parameters somehow related to the process of speech production Some of these model only a class of speech signals eg vowel modeling using extracted formants

As shall be seen in the following section most of the phoneme utterances are chaotic in nature in the sense that they have positive second-order entropy Similar analysis was done on stationary interval segments of an uttered signal [43] This means that the vocal system is a non-linear system because only such systems are capable of displaying chaotic behaviour In this sense, modeling speech signals using the constraint of linearity may be far from optimum The following section discusses some practical aspects of finding the correlation dimension and second-order entropy from observed data

3.3 Estimation of Dimensions and Entropies from Scalar Time Series : Practical Aspects and Results

As mentioned in the previous section, some utterances eg fricatives are produced by a turbulent flow of air through a narrow constriction These are modeled using random excitation, while in other utterances, eg vowels, one searches for regular behaviour One of the offshoots of the study of chaotic behaviour has been the proposal of a variety of dimensions that can be used to

get an idea of the minimum number of independent variables needed to model observed behaviour For example, if the dimension is found to be 2 23, then the minimum number of independent variables needed is the next higher integer which in this case is 3 If the dimension is found to be large, then the number of independent variables needed to model the system would also be large In such situations stochastic modeling schemes may be preferred Similarly, the entropy of a particular trajectory gives an idea of its nature eg for regular behaviour, the Kolmogorov entropy $K_1 = 0$ For chaotic behaviour, $K_1 > 0$, while $K_1 \rightarrow \infty$ signifies random behaviour The inverse of the entropy gives an idea of the length of time upto which predictions can be carried out Thus, for regular behaviour, one can predict for infinite time while for random behaviour, the prediction time tends to zero

Amongst the three dimensions and the corresponding entropies presented in Chapter 2, the correlation dimension D_2, and the second-order entropy K_2 are most frequently used because of their ease of computation They use the attractor reconstruction theorems presented in Chapter 2 The theorems ensure that it is only necessary to observe one variable of the system evolving in time It is not necessary to measure all the n variables, y_k, k=0,1, ,n-1 In fact, it is also not necessary to know the value n (i e , the dimension of the real space) to estimate the attractor dimension The measurement of a single variable y_i, i=0,1, is sufficient because it contains all the relevant information about the other (n-1) variables in $d^k y/dt^k$

An analytical procedure that required the knowledge of all the n variables would be impractical However, the underlying idea is to reconstruct a d-dimensional vector-time series X_i, i=0, 1, from the observed variable y_i, i= 0, 1, using delays as follows

$$X_0 = [\, y_0 \; y_m \; y_{2m} \quad\quad y_{(d-1)\,m} \,]^T$$

$$X_1 = [\, y_1 \; y_{1+m} \; y_{1+2m} \quad\quad y_{1+(d-1)\,m} \,]^T$$

$$X_1 = [\, y_1 \; y_{1+m} \; y_{1+2m} \quad\quad y_{1+(d-1)\,m} \,]^T \tag{3 3 1}$$

There is no hard and fast rule for selecting the value of the delay m in equation (3 3 1) above Some guidelines have been proposed which are in some sense optimal The central idea in choosing the proper delay is that the coordinates of the reconstructed state-space be as uncorrelated as possible A simple procedure to choose m would be to look for the decay or the first zero in the autocorrelation function $\langle y_i \, y_{i+m} \rangle$ This is often unacceptable because the time series may have periodicities such that the autocorrelation function oscillates for quite a few delays Also zero autocorrelation means only second-order uncorrelatedness Another procedure to find m is to use the delay corresponding to the first minima of mutual information [56,58]

For information theory to apply, the probabilities of messages considered must exist The application of information theory to strange attractors is justified because they are ergodic and have well-defined asymptotic probability distributions Thus, the probabilities of the messages exist and long time averages converge to probabilites The following discussion briefly reviews the mutual information concept as applicable to the problem of state-space reconstruction

Consider a process in which messages are being sent Let S denote the system which consists of a set of possible messages s_1, s_2, , s_n and associated probabilities $P_s(S_1)$, $P_s(S_2)$, , $P_s(S_n)$ The average amount of information gained from a measurement that specifies s is the *entropy* H of a system,

$$H(S) = -\sum_{1} P_S(s_1) \log P_S(s_1) \qquad (3\,3\,2)$$

In the present problem, one is interested in measuring how dependent the values of y_{n+k} are on y_n. By making the assignment $[s,q] = [y_n, y_{n+k}]$, one can consider the general coupled system (S,Q). Then $H(Q/s_1)$ denotes the uncertainty in the measurement of q given that s was found to be s_1. $H(Q/s_1)$ is given by

$$H(Q/s_1) = -\sum_{J} P_{q/s}(q_J/s_1) \log [P_{q/s}(q_J/s_1)] \qquad (3\,3\,3)$$

$$= -\sum_{J} [P_{sq}(s_1,q_J)/P_S(s_1)] \log [P_{sq}(s_1,q_J)/P_S(s_1)] \qquad (3\,3\,4)$$

where $P_{q/s}(q_J/s_1)$ is the probability that a measurement of q will yield q_J given that the measured value of s is s_1.

Similarly, $H(Q/S)$ denotes the average uncertainty in the measurement of y_{n+k} given that y_k has been measured. It is given by

$$H(Q/S) = \sum_{1} P_S(s_1) H(Q/s_1)$$

$$= -\sum_{1,J} P_{sq}(s_1,q_J) \log [P_{sq}(s_1, q_J) / P_S(s_1)]$$

$$= H(S,Q) - H(S), \qquad (3\,3\,5)$$

where

$$H(S,Q) = -\sum_{1,J} P_{s,q}(s_1,q_J) \log [P_{sq}(s_1,q_J)] \qquad (3\,3\,6)$$

$H(Q)$ denotes the uncertainty of q in isolation and $H(Q/S)$, the uncertainty of q given a measurement of s. So the amount that a measurement of s reduces the uncertainty of q is

$$I(Q,S) = H(Q) - H(Q/S)$$

$$= H(Q) + H(S) - H(S,Q)$$

$$= I(S,Q) \qquad (3\,3\,7)$$

This is called the *mutual information*. If all the logs are to the base 2, then H is in units of bits. Mutual information provides an answer to how many bits

on the average can be predicted about q given a measurement of s Finding the value of k at which minimum mutual information of the system (S,Q) is observed, is in a way finding the delay k at which the values of k are least dependent

The results of the application of this technique to phoneme time series will be presented after a description of the experimental set-up

The experimental set-up consisted of an 8-bit analog to digital converter which sampled the intensity of all phoneme utterances (in our case the phonemes of the IPA) at 20 0 kHz and stored them in different files The data from each of these phoneme time series was used for the diagrams and calculations

Figures 3 2 - 3 10, part (c) show the reconstructed phase space trajectories y_{n+k} vs y_n for phonemes /u/, /ɑI/, /r/, /m/, /f/, /v/, /g/, /k/ and /ʧ/ at k=2 Successive points on the trajectory have been joined by straight lines assuming that the acutal trajectories can be approximated thus Figure 3 11 shows the trajectories of vowels /u/, /U/, /o/, /ɔ/ and /ɑ/ using k=2 It shows the slow change in overall shape from one phoneme trajectory to the next

The first minima of the mutual information was calculated for all phoneme time series to obtain the best delay k For 30 phonemes, this was observed at k=1, for 9 phonemes it was observed at k=2 while for the remaining it was observed at k=4

As discussed in Chapter2, the correlation or second-order dimension D_2 is given by

$$D_2 = \lim_{r \to 0} \frac{\log C \ (r)}{\log r},$$
(3 3 8)

where

$$C \ (r) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=1}^{N} \theta \ (\ r - |X_i - X_j| \ )$$
(3 3 9)

θ is the Heaviside function $\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$, otherwise The
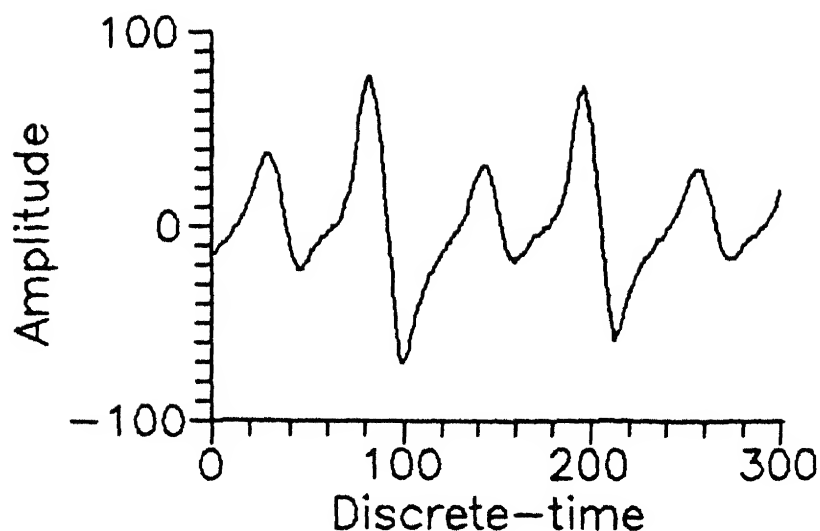
a)
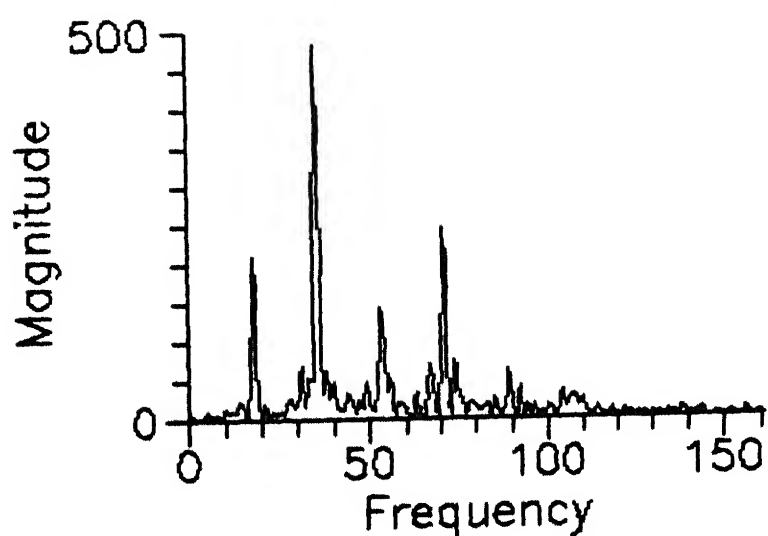
b)
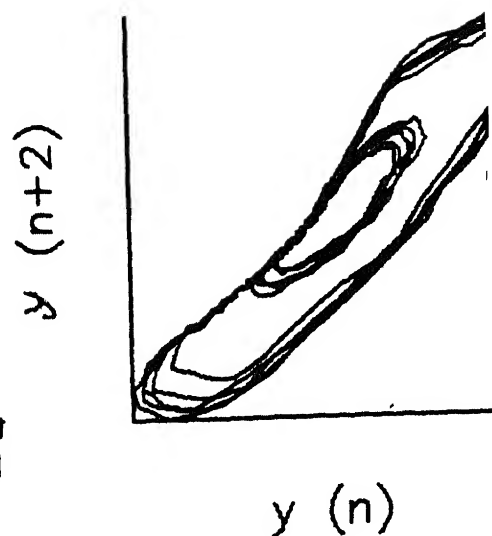
c)

Figure 3.2 - All graphs for phoneme /u/, type - *vowel*, sampled at 20 0 kHz using 8-bit ADC. a) Part of the time-series, b) First 150 points of a 1024-point DFT, c) Phase-space plot using 600 points. In all the three cases, adjacent points have been joined by straight lines.

a)
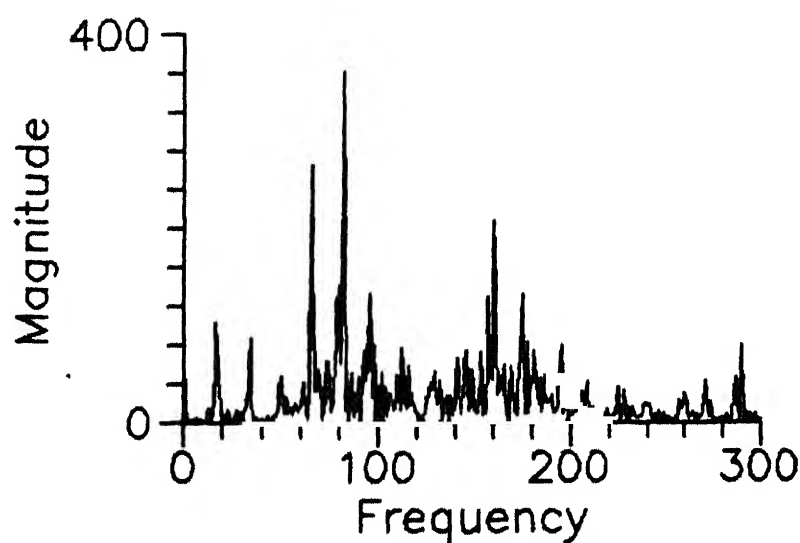
b)

c)

Figure 3.3 - All graphs for phoneme /aI/, type - *dipthong*, sampled at 20.
kHz. a) Part of the time-series, b) First 300 points of a 1024-point DFT of the
phoneme, c) Phase-space plot using 600 points

a)
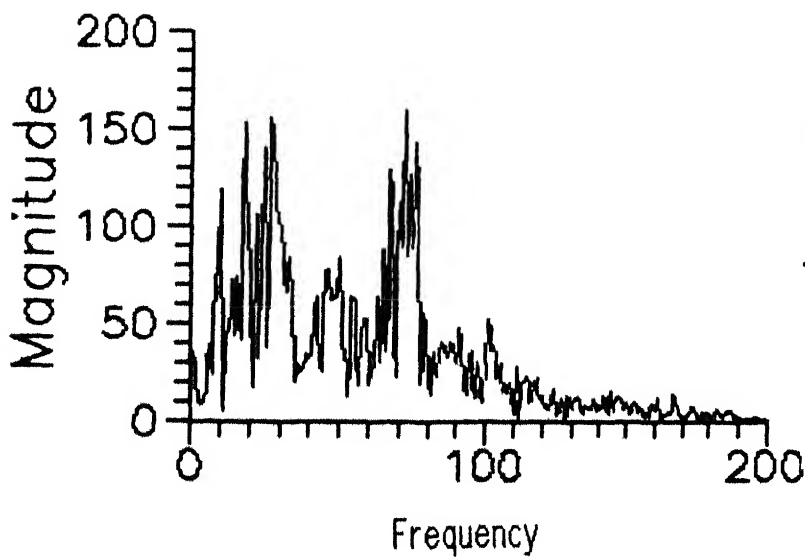
b)
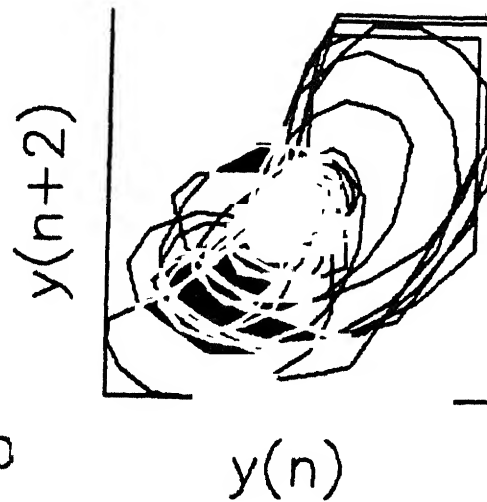
c)

Figure 3.4 - All graphs for phoneme /r/, type - *semi-vowel*, sampled at 20.0 kHz a) Part of the time-series b) First 200 points of a 1024-point DFT, c) Phase-space plot using 900 points

a)

Frequency
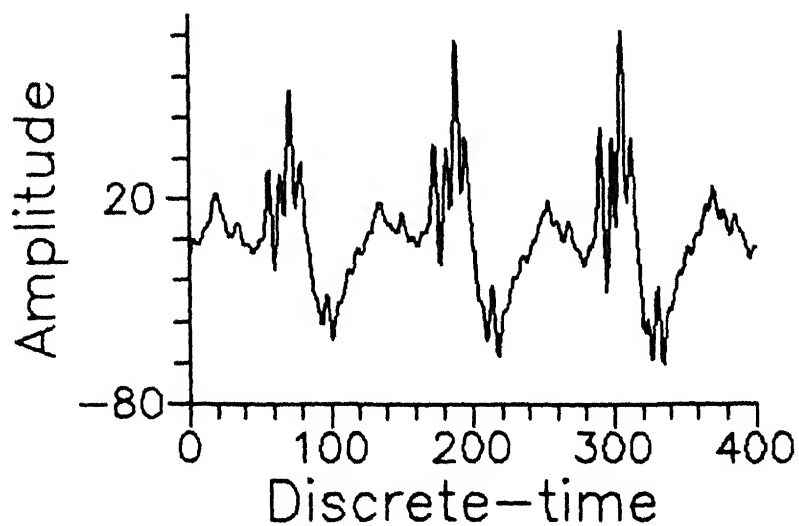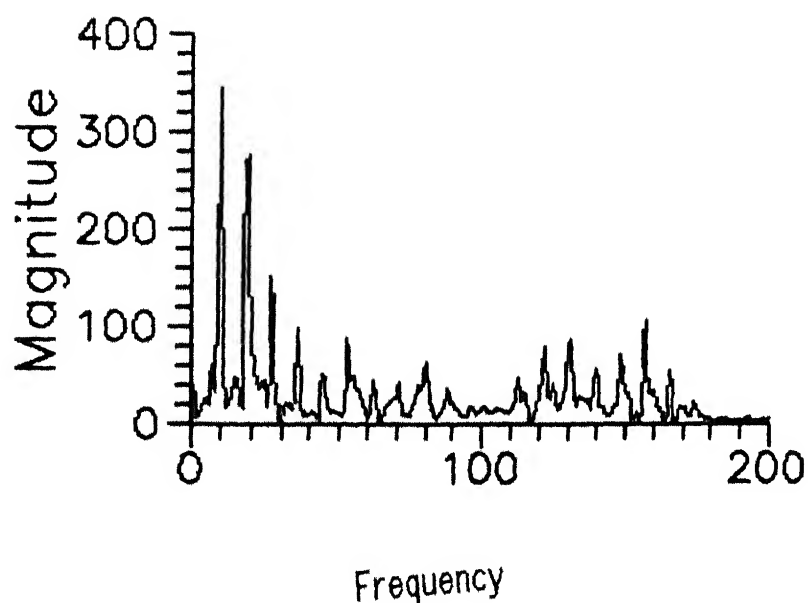
b)
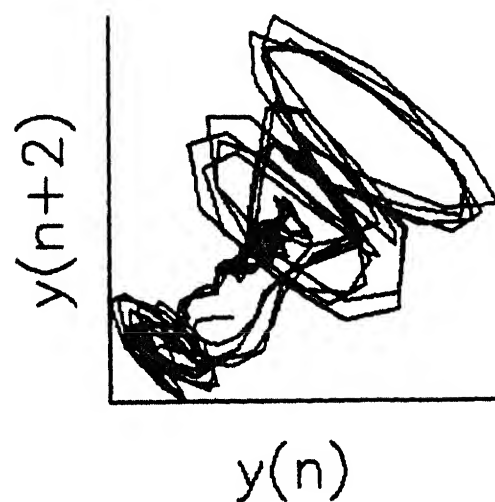
c)

Figure 3.5 - All graphs for phoneme /m/, type - *nasal*, sampled at 20.0 kHz a) Part of the time-series b) First 200 points of a 1024-point DFT, c) Phase-space plot using 500 points of the time-series
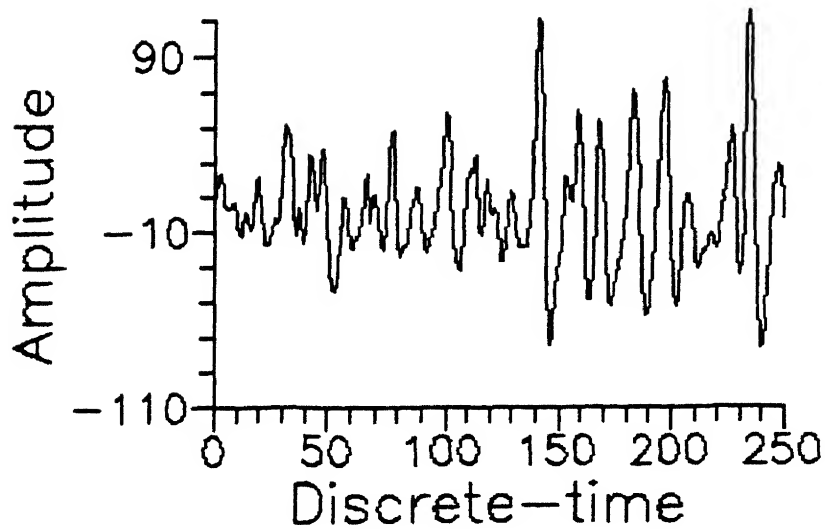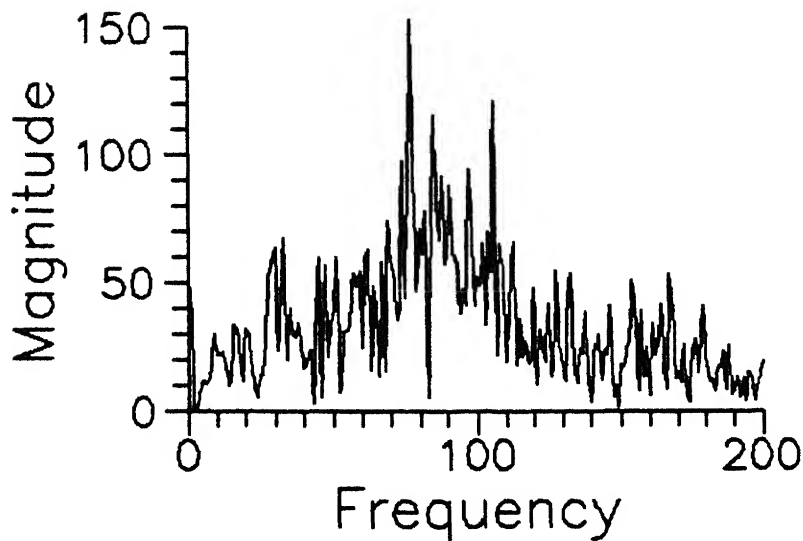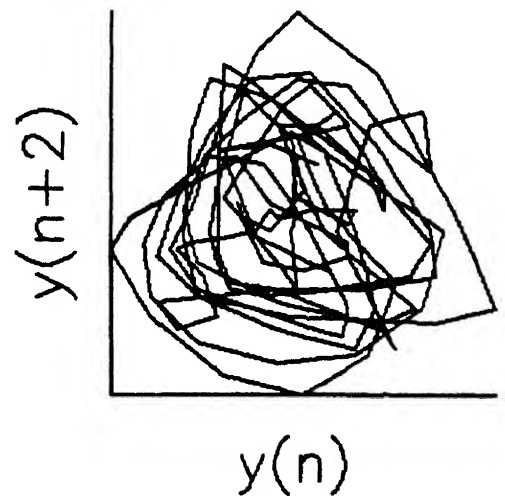
Figure 3.6 - All graphs for phoneme /f/, type - *unvoiced fricative*, sampled at 20 0 kHz. a) Part of the time-series, b) First 200 points of a 1024-point DFT, c) Phase-space plot using 250 points
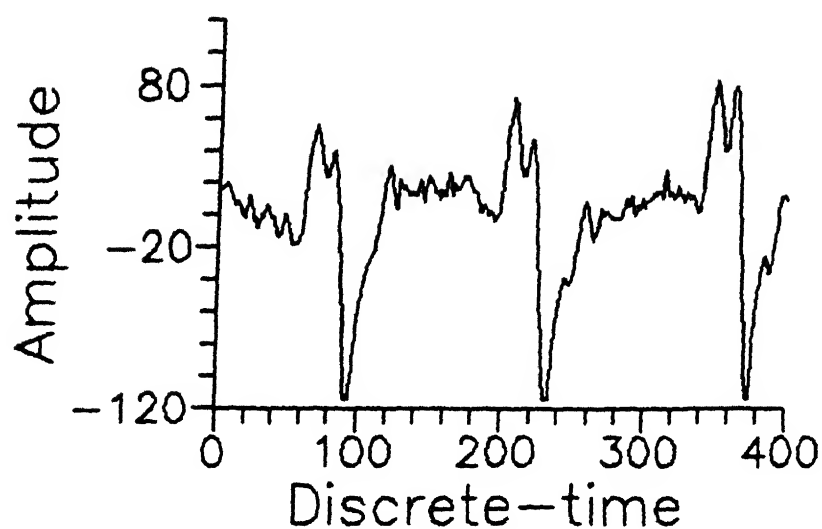
a)
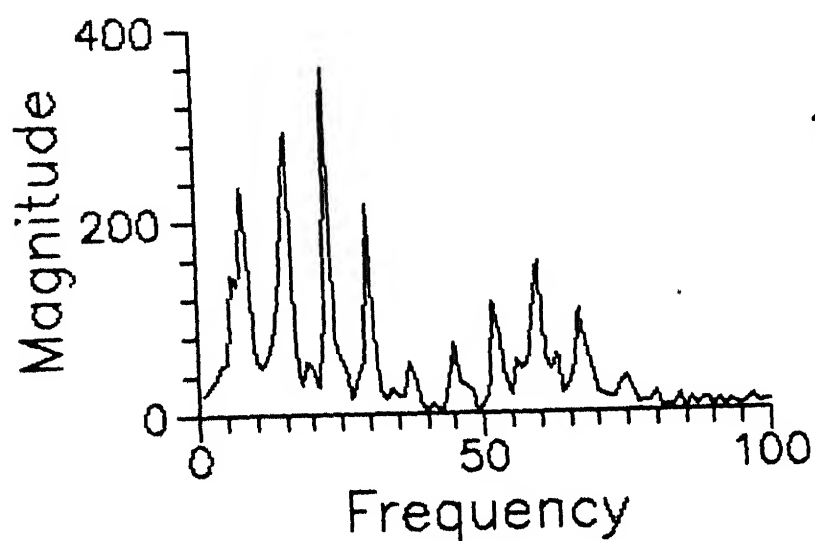
b)
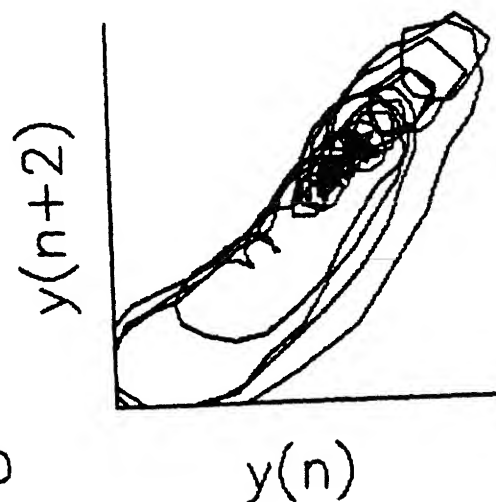
c)

Figure 3.7 - All graphs for phoneme /v/, type - *voiced fricative*, sampled at 20.0 kHz. a) Part of the time-series, b) First 100 points of a 1024-point DFT, c) Phase-space plot using 500 points
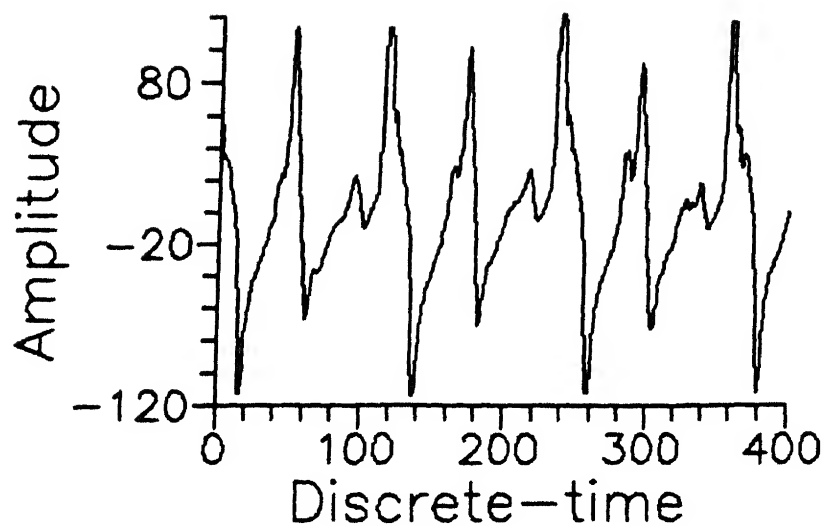
a)
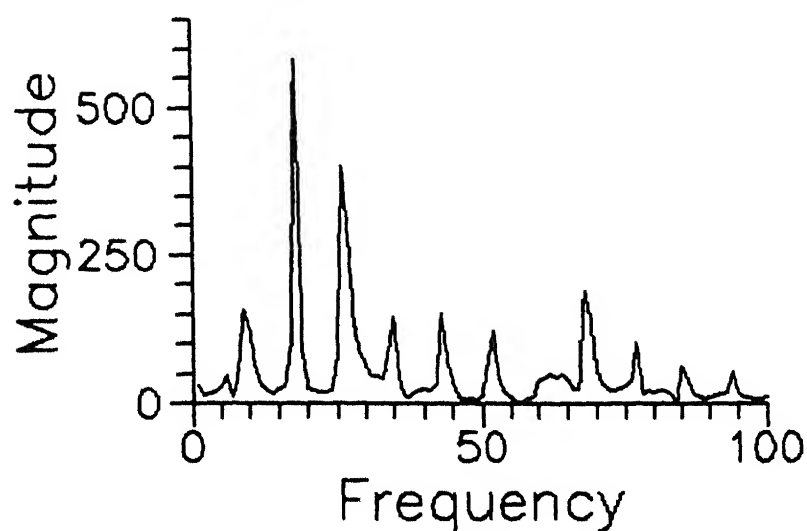
b)
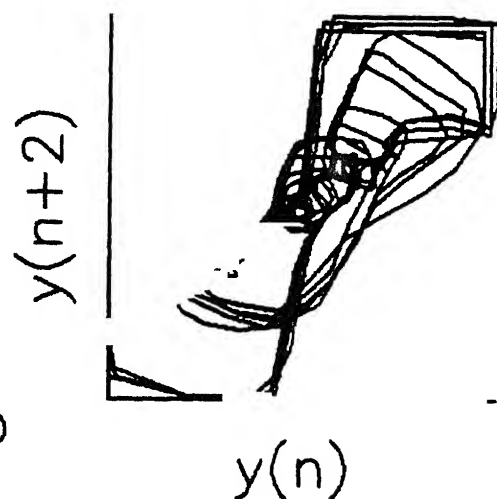
c)

Figure 3.8 - All graphs for phoneme /g/, type - *voiced stop*, sampled at 20 0 kHz a) Part of the time-series, b) First 100 points of a 1024-point DFT, c) Phase-space plot using 600 points
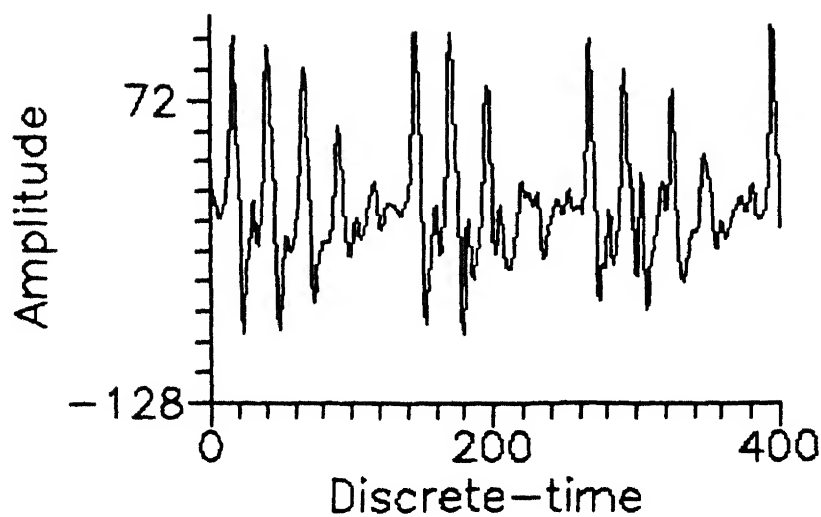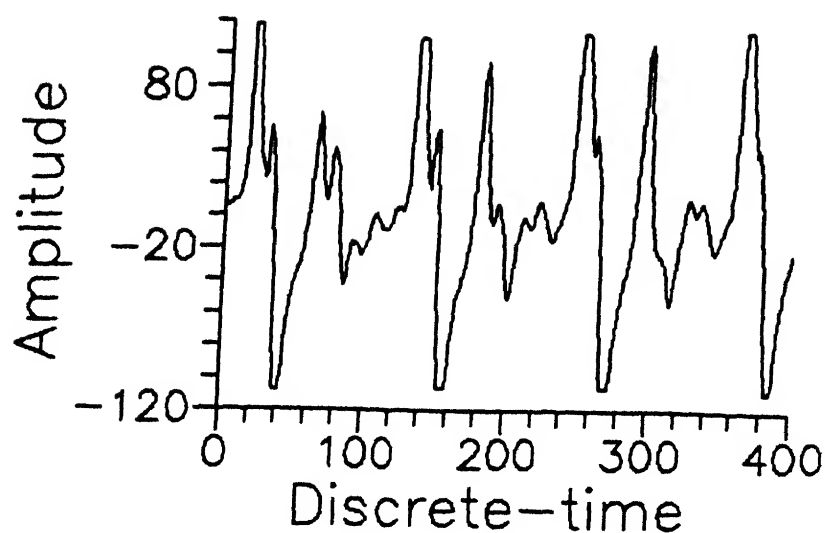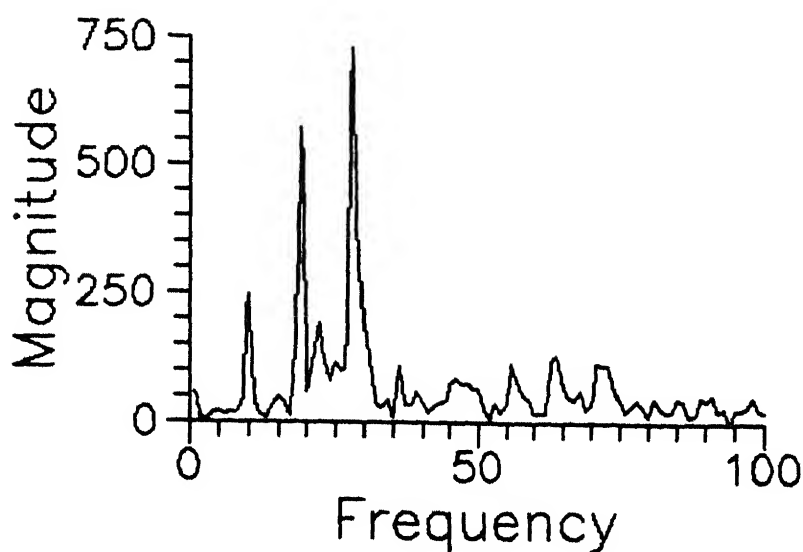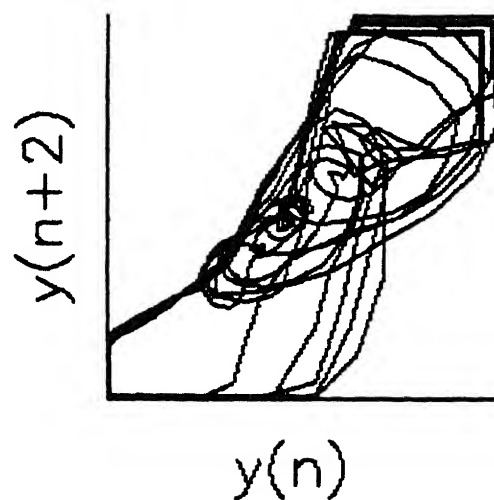
Figure 3.9 - All graphs for phoneme /k/, type - *unvoiced stop*, sampled at 20 0 kHz a) Part of the time-series, b) First 100 points of a 1024-point DFT, c) Phase-space plot using 600 points
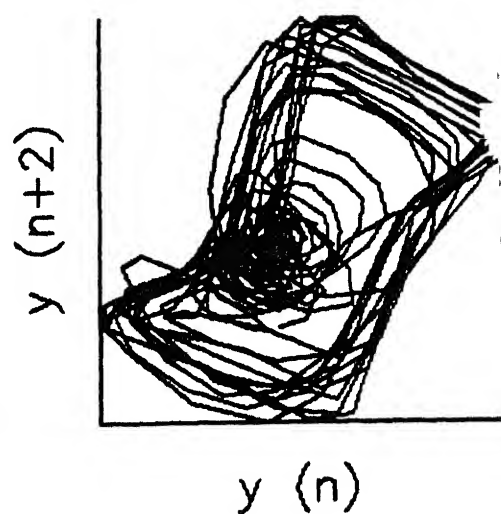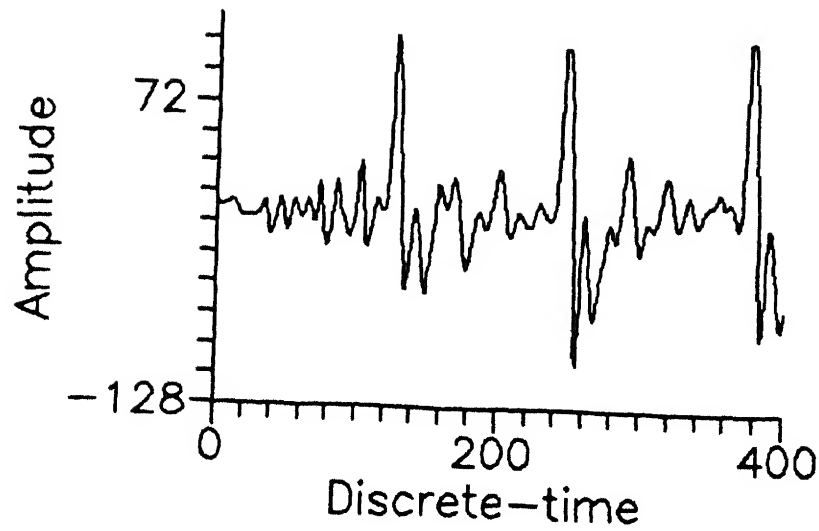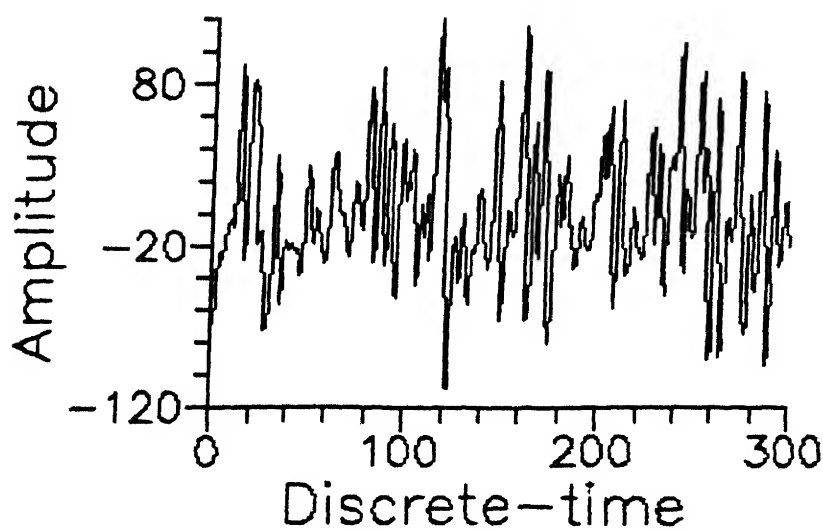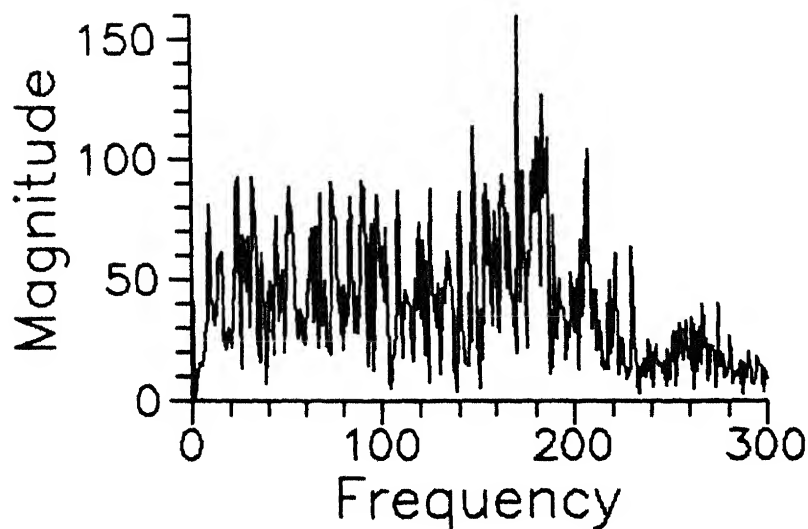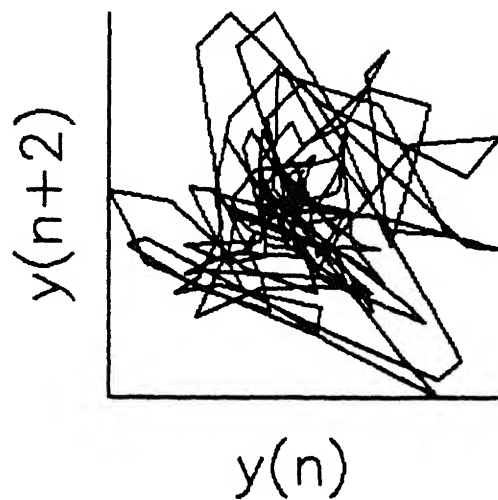
Figure 3.10 - All graphs for phoneme /tʃ/, type - *unvoiced affricate*, sampled at 20.0 kHz a) Part of the time-series, b) First 300 points of a 1024-point DFT, c) Phase-space plot using 250 points
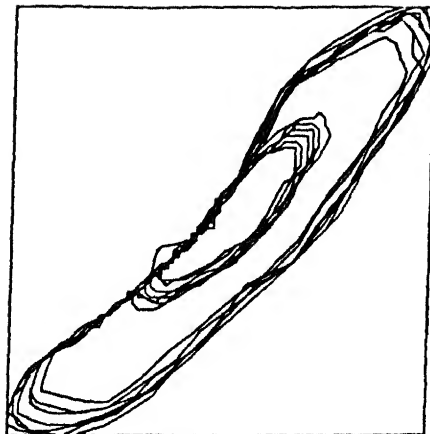
a)
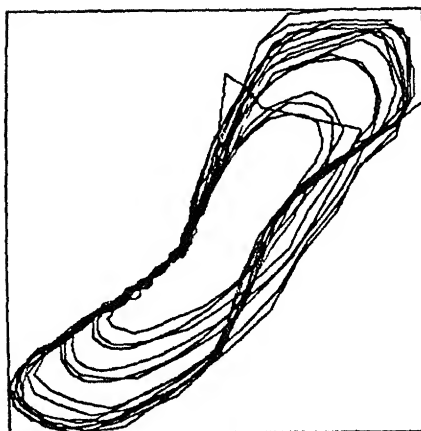
b)
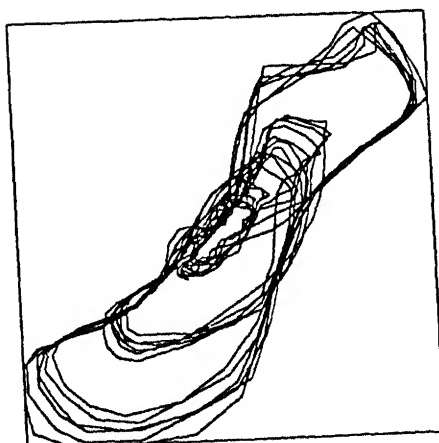
c)
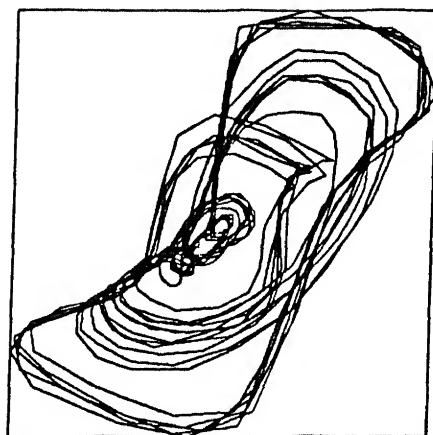
d)
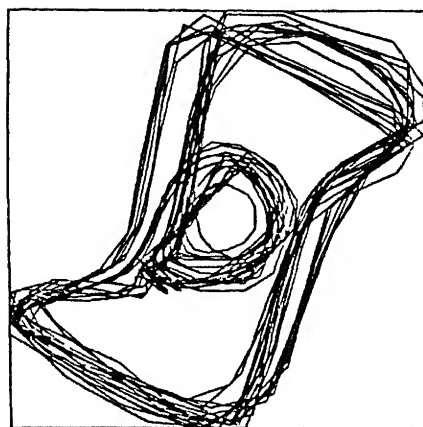
e)

Figure 3 11 - Phase-space plots y(n+k) vs. y(n) using k=2 for various vowels a) Phoneme /u/, b) Phoneme /U/, c) Phoneme /o/, d) Phoneme /ɔ/ and e) Phoneme /a/ Successive points on the trajectory have been joined by straight lines assuming that the actual trajectories can be approximated thus The overall shape of the trajectories change slowly in the order of phonemes above

function C (r) counts the number of pairs of those points with a distance $|X_1 - X_j|$ less than r

In practice, the dimension d of the reconstructed vector time series X_1, is increased from 1 For each value of d, X_1 is used to find C (r) for decreasing r The hypothesis in finding the correlation dimension is that C (r) scales as a power law with r i e ,

$$C (r) \propto r^{D_2} \qquad\qquad (3\ 3\ 10)$$

Takens' theorems can be used to show that if d is sufficiently large, then the correlation dimension D_2 of the original system (i e , the actual n-variable system from which a single variable is measured) and the reconstructed d-dimensional system will be the same As d is increased from 1, D_2 will also increase until it reaches its correct value Increases in d beyond this point will not affect D_2 beyond errors, although the calculations become increasingly slow as d is increased Then we plot the correlation dimension versus the reconstruction or embedding dimension If the curve keeps growing (i e , D_2 continues to increase with d) then the system approaches truly random behaviour with the limits of experimental observation

This procedure even provides a scheme for detecting a low-dimensional attractor that is contaminated by noise [26] If there is noise of magnitude ξ added to deterministic data, a plot of log C (r) versus log r will have a knee at ξ Figure 3 12 (a) shows the log C(r) vs log r plot for increasing d (i e , d=1, 2, , 18) for the phoneme /θ/ The slope above the knee will give the correct dimension for the deterministic system which in this case is the vocal system The slope below the knee will give the dimension corresponding to the noise Figure 3 12(b) shows the plot D_2 vs d for the same phoneme The saturated curve shows that D_2 for the vocal tract configuration during the utterance of the phoneme /θ/ is 2 17±

0 07

Similar to the correlation dimension D_2, the second-order entropy can be defined in terms of the correlation function C (r) as follows

$$K_2 = \lim_{r \to 0} \lim_{d \to \infty} \frac{1}{\tau} \log \frac{C_d \ (r)}{C_{d+1} \ (r)} \tag{3 3 11}$$

where C_d (r) is C (r) given by equation (3 3 9) for the embedding dimension d, and τ is the time difference between succesive points on the reconstructed vector time series which in our case is 0 05 ms Thus, K_2 can be estimated from the vertical distances (at the same values of r) between the curves belonging to successive dimensions d in the log C(r) vs log r plots for increasing d (eg Figure 3 12(a))

Figure 3 12(c) shows the plot between $\tau^{-1} \log [C_d \ (r) / C_{d+1} \ (r)]$ for increasing d for the phoneme /θ/ The indicated values for each particular d was calculated from the mean value of C_d (r)/ C_{d+1} (r) over the linear range of r above the knee (ie, above the noise amplitude) from figure 3 12(a) The limiting value of $\tau^{-1} \log [C_d(r)/C_{d+1}(r)]$ for large d gives the second-order entropy K_2 which in this case is $\, 6082 \pm 312$

Table 3 1 shows the correlation dimension D_2 while Table 3 2 shows the second-order entropy K_2 for all phonemes of the IPA Further experimental details and results are discussed below

1 Because of the constraint that the vocal tract changes configuration for non-continuant phonemes and even otherwise minor changes occur, we had to base our calculation on N=500 data points For the calculation of the correlation function, a large value of N is required Therefore, we sampled the phonemes at a relatively large sampling rate of 20 0 kHz to allow us to use 500 data points assuming that the vocal tract configuration does not change for 25 ms

2 The log C(r) vs log r plots for different embedding dimensions D have a sufficiently large linear range to arrive at conclusive results The dimensions and entropies given in Table 3 1 and 3 2 respectively, are not contaminated by low

a)
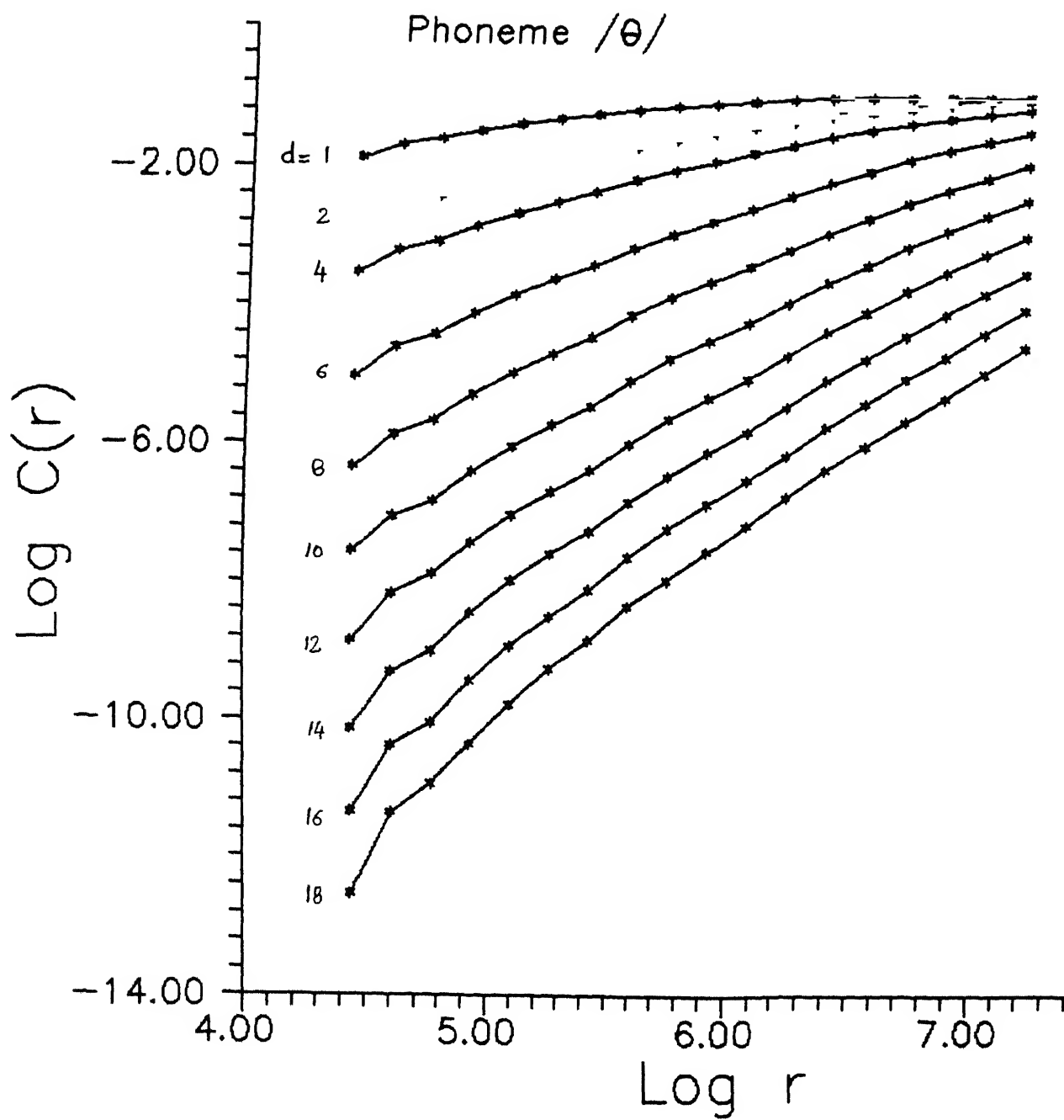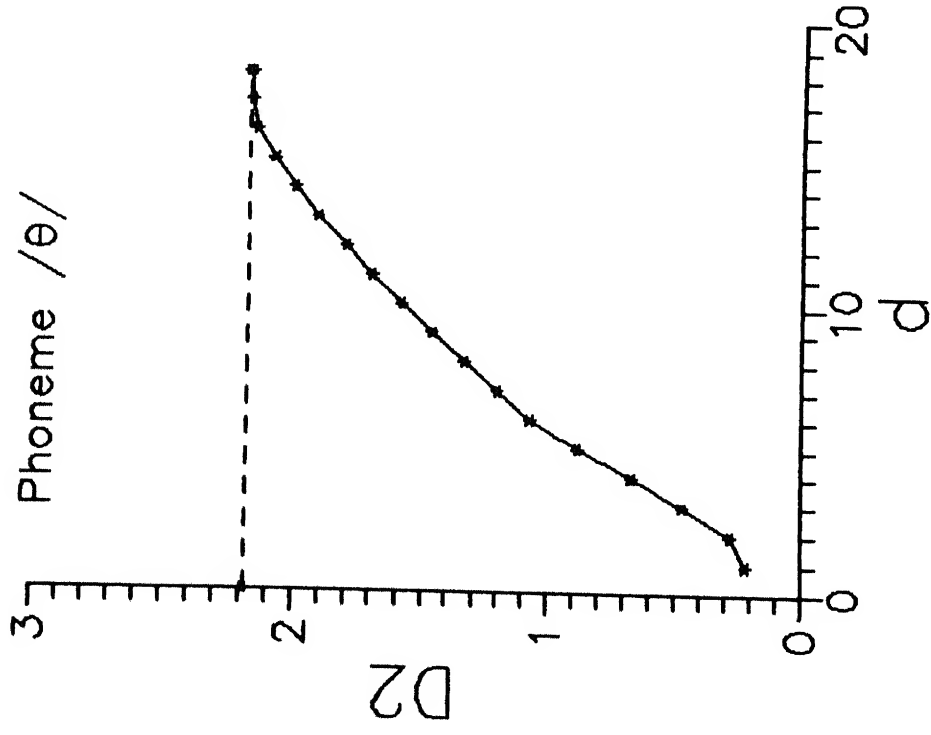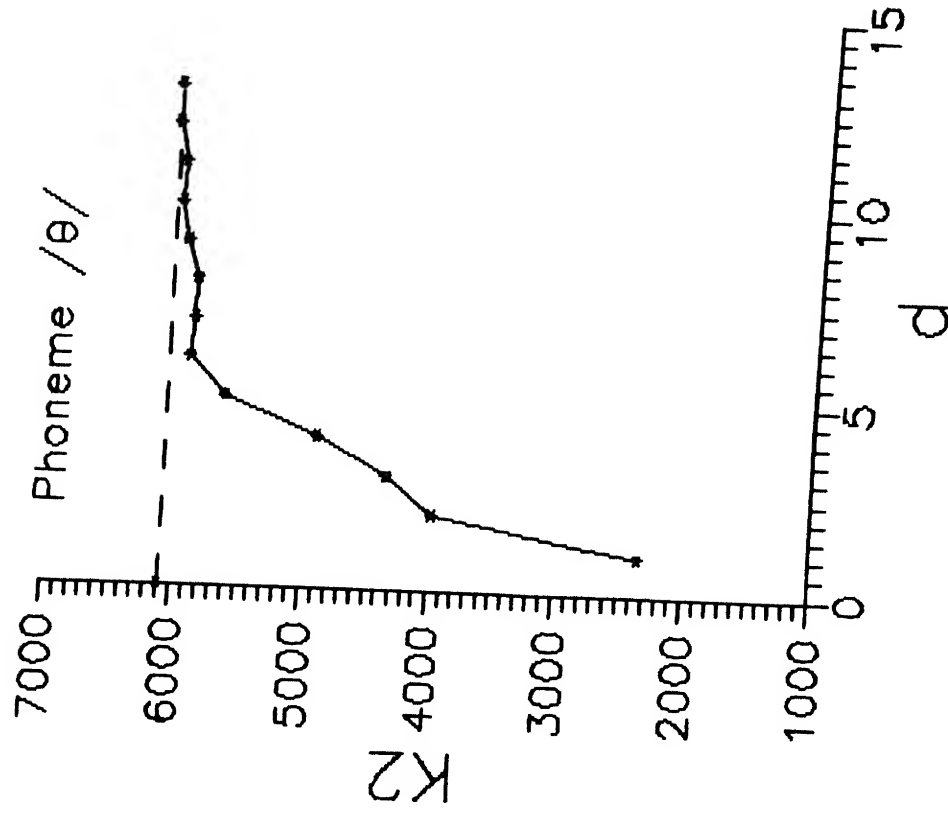
Figure 3.12 - All graphs for phoneme / / a) Log C(r) vs log r for embedding
dimension d = 1,2,4,6,8,10,12,14,16,18 Note the increase in the slope at low
values of r compared to higher r for same d, showing the presence of low
amplitude noise in the analysed data, b) Plot of the slope of log C(r) vs. log r
for various d The correlation dimension obtained is 2.17 ± 0.07 c) Plot of
entropy vs varying d The value of entropy obtained from the saturated curve
is 6080 - 310

Phoneme	Type	Correlation Dimension
/ɪ/	Vowel-Front	3 10 ± 0 19
/I/	-do-	2 23 ± 0 41
/e/	-do-	2 50 ± 0 10
/ɜ/	-do-	2 74 ± 0 28
/ae/	-do-	2 00 ± 0 23
/a/	-do-	2 60 ± 0 10
/u/	Vowel-Back	2 00 ± 0 10
/U/	-do-	1 74 ± 0 13
/o/	-do-	1 70 ± 0 18
/ɔ/	Vowel - Mid	1 88 ± 0 05
/ɑ/	-do-	2 06 ± 0 04
/ʌ/	-do-	2 07 ± 0 14
/ə/	-do-	2 20 ± 0 10
/ɜ/	-do-	2 00 ± 0 20
/ɚ/	-do-	2 42 ± 0 08
/w/	Semivowel-Liquid	4 44 ± 0 14
/ʍ/	-do-	6 70 ± 0 07
/l/	-do-	1 86 ± 0 01
/r/	Semivowel-Glide	1 42 ± 0 24
/j/	-do-	3 23 ± 0 21
/aɪ/	Dipthong	3 40 ± 0 08
/aU/	-do-	2 11 ± 0 21
/ɔI/	-do-	2 70 ± 0 10
/b/	Voiced Stop	2 90 ± 0 45
/d/	-do-	2 50 ± 0 64
/g/	-do-	2 56 ± 0 12
/p/	Unvoiced Stop	2 28 ± 0 64
/t/	-do-	2 90 ± 0 19
/k/	-do-	2 69 ± 0 39
/v/	Voiced Fricative	5 70 ± 0 45
/z/	-do-	4 99 ± 0 31
/ð/	-do-	2 65 ± 0 08
/ʒ/	-do-	3 17 ± 0 29
/f/	Unvoiced Fricative	5 80 ± 0 45
/s/	-do-	5 50 ± 0 27
/θ/	-do-	2 17 ± 0 07
/ʃ/	-do-	7 90 ± 0 40
/dʒ/	Affricate	1 64 ± 0 18
/tʃ/	-do-	6 30 ± 0 10
/h/	Whisper	5 79 ± 0 28
/m/	Nasal	4 11 ± 0 70
/n/	-do-	2 72 ± 0 37
/ŋ/	-do-	2 76 ± 0 20

TABLE 3.1 - Correlation dimension D_2 for all phonemes of the IPA The phoneme utterances were sampled at 20 kHz using 8-bit ADC N=500 data points was used The error ranges were calculated from the extrapolation of the worst possible slopes of the linear range of log C(r) vs log r curves for each phoneme

Phoneme	Type	Second-Order Entropy
/ɪ/	Vowel-Front	7100 ± 2064
/I/	-do-	5140 ± 1712
/e/	-do-	6260 ± 290
/ɜ/	-do-	8480 ± 1768
/ae/	-do-	4729 ± 968
/a/	-do-	5230 ± 1480
/u/	Vowel-Back	2600 ± 1084
/U/	-do-	3898 ± 688
/o/	-do-	1804 ± 48
/ɔ/	Vowel – Mid	1280 ± 578
/ɒ/	-do-	1760 ± 1960
/ʌ/	-do-	6740 ± 392
/ə/	-do-	6370 ± 400
/ɝ/	-do-	6005 ± 2124
/ɚ/	-do-	4830 ± 1824
/w/	Semivowel-Liquid	9446 ± 3316
/ʍ/	-do-	6474 ± 480
/l/	-do-	5010 ± 768
/r/	Semivowel-Glide	790 ± 1252
/j/	-do-	6310 ± 888
/aɪ/	Dipthong	7110 ± 2496
/aU/	-do-	7840 ± 1840
/ɔɪ/	-do-	5801 ± 824
/b/	Voiced Stop	8665 ± 2116
/d/	-do-	7100 ± 880
/ɡ/	-do-	6228 ± 1732
/p/	Unvoiced Stop	480 ± 3776
/t/	-do-	7310 ± 2212
/k/	-do-	10300 ± 1328
/v/	Voiced Fricative	11910 ± 8100
/z/	-do-	12200 ± 1152
/ð/	-do-	7400 ± 2112
/ʒ/	-do-	7406 ± 1800
/f/	Unvoiced Fricative	✳
/s/	-do-	36735 ± 130
/θ/	-do-	6082 ± 312
/ʃ/	-do-	25995 ± 1310
/dʒ/	Affricate	600 ± 2240
/tʃ/	-do-	✳
/h/	Whisper	7362 ± 5020
/m/	Nasal	7717 ± 716
/n/	-do-	7160 ± 180
/ŋ/	-do-	5906 ± 400

TABLE 3.2 - The second-order entropy K_2 for all phonemes of the IPA uttered by an adult male N=500 points were used for the calculations The error ranges correspond to the worst possible slopes of the linear range of the log C(r) vs log r plots

amplitude noise The presence of low amplitude noise was distinguished from the actual data in the log C(r) vs log r plots as discussed earlier in the section The error bounds given for the estimated dimensions and entropies are based on the worst slopes from the linear range of the log C(r) vs log r plots for different d

3 Simulations done for N = 1000 and 1500 on sustained utterances of some of the continuant phonemes showed very little variation of the dimension

4 Most of the attractors of vowel time series have dimension less than 3 0 Consonants were generally found to have higher dimension than vowels This shows the more complex behaviour of consonants particularly fricatives as can also be seen from a comparison of phase space plots of fricatives with those of vowels Studies of utterances of 3 adult males showed that the dimensions of the unvoiced phonemes and their voiced counterparts are nearly the same Thus, the role of excitation is not brought out by the dimension invariant, assuming that the vocal tract configuration remains the same for both

5 A look at the second-order entropy K_2 given in Table 3 2 shows that for almost all phonemes, it is positive within the error range Since $K_1 > 0$ is a sufficient condition for chaos and $K_2 \leq K_1$, this means that most phoneme time series are chaotic in nature For almost all vowels, K_2 lies in the range 1000 s^{-1} to 8000 s^{-1}, while for most consonants, it lies in the range 6000 s^{-1} to 25 000 s^{-1} and is concentrated in the region of 7000 s^{-1} to 12 000 s^{-1}

3.4 Concluding Remarks

The main point of this chapter was to present the results of the dimensional analysis of attractors and entropies of the trajectories reconstructed from all phoneme utterances of the IPA This was done after an introduction of the vocal tract which is the underlying dynamical system, and all the classes of phonemes which form the basis of our analysis throughout the thesis It was also pointed out that a model of the vocal tract that incorporated

all the observed effects is not yet available. The fact that we do not know the dimension of the state space in which the vocal system actually resides does not hamper our investigation of the dimensions of the attractors in which the trajectories eventually reside. This is because of the powerful theorems of Takens discussed in the previous chapter. Finally, analysis showed that the underlying attractors of phonemes are low dimensional. This means that a few variables are needed to model the utterances deterministically using non-linear functional representations and makes a strong case for their study. The next chapter discusses the signal modeling problem in this framework.

CHAPTER 4

SIGNAL MODELING : THEORY AND RESULTS OF APPLICATION TO SPEECH SIGNALS

As mentioned in Chapter 1, it is now known that apparently random behaviour can arise from simple non-linear systems Until recently, the only tool to analyse such behaviour was based on Kolmogorov's theory of random processes Alternatively, one may postulate that a time series is the evolution of a dynamical system because most natural phenomena can be modeled as such Coupled with this the assumption that randomness arises out of chaos rather than intractable complexity leads to a fundamentally different approach to signal modeling which will be the main theme of this chapter In Chapter 1, a brief review of the auto-regressive and state-space modeling schemes was presented Signal modeling in the present framework differs from autoregressive schemes in two main respects Firstly, there is no assumption of a stationary random process generating the time series and secondly, the present modeling scheme uses non-linear functional representations as opposed to the linear ones in the auto-regressive schemes The present modeling scheme also differs from the deterministic state-space modeling schemes in the following respects Firstly, there is no attempt to linearize trajectories using the concept of nominal trajectories because complex behaviour can arise out of simple non-linear systems which one tries to model in the present framework Secondly, the traditional state-space modeling schemes require the knowledge of the variables involved or the state-space itself whereas in the present scheme, the observation of a single variable allows us to reconstruct the state-space

The chapter begins with the theoretical framework for the signal modeling problem Then some practical aspects of the problem are discussed Next a

discussion on the various types of functional representations is given followed by the two approximation techniques - global and local which are used to obtain the parameters of the model The local approximation technique is not practical from the storage and transmission point of view because of the requirement of a prohibitively large model order but it has better prediction properties than the global approximation technique We, thus, propose and study a compromised overlapping neighbourhood (CON)-local approximation technique that reduces the model order and yet tries to retain the better prediction properties of the local approximation technique The prediction properties of the global and CON-local approximation technique are compared with the LPC using speech data in the form of phoneme utterances Finally, the numerical complexity of the two approximation techniques is compared with the LPC

4.1 The Theoretical Framework

Let x_i, $i = 1, 2,$ be a scalar time series evolving in time Consider it to be the observed variable of a dynamical system with a map or vector field φ $M \rightarrow$ M, M being a compact manifold on which the phase space of the system is located That is, there exists an *observable* on the dynamical system which is a smooth function y $M \rightarrow \mathbb{R}$ such that $x_i = y (z_i)$, $1 \leq i < \infty$, where z_i is a point on the manifold corresponding to which x_i is the observed variable Use $\Phi_{(\varphi,y)}$ $M \rightarrow \mathbb{R}^k$ as follows

$$\Phi_{(\varphi,y)} (z_i) = (y(z_i), y(\varphi(z_i)), y(\varphi^2(z_i)), \quad , y(\varphi^{k-1}(z_i)))\qquad(4 1 1)$$

to embed the attractor in real space $\mathbb{R}^k$ Takens' theorem requires for sufficiency that $k \geq 2m + 1$, where m is the dimension of the manifold M Let f $\mathbb{R}^k \rightarrow \mathbb{R}^k$ be a smooth map such that $\Phi_{(\varphi,y)} (z_i) = f^i \Phi_{(\varphi,y)} (z_0)$, $1 \leq i < \infty$ describes the embedded trajectory in $\mathbb{R}^k$ The signal modeling problem then is to construct a smooth map $\tilde{f}_N \cdot \mathbb{R}^k \rightarrow \mathbb{R}^k$ using only a finite number of iterates z_i, $1 \leq i \leq N$, for which

$$\Phi_{(\varphi,y)}(z_{i+1}) = \widetilde{f}_N \, \Phi_{(\varphi,y)}(z_i), \quad 1 \leq i \leq N-1$$

A simplification of notation is now in order. As the signal modeling problem deals solely with the embedded trajectory in $\mathbf{R}^k$, $\Phi_{(\varphi,y)}(z_i)$ will henceforth be replaced by X_i which is a vector of the k delayed elements of the observed time series x_i, $i = 1, 2, \ldots$ i.e.,

$$X_i = [\ x_i \quad x_{i+1} \quad \ldots \quad x_{i+k-1}\ ]^T \tag{4 1 2}$$

For $k = 1$, the signal modeling problem reduces to finding a good approximation to f. For $k > 1$, this becomes a problem of geometrically fitting k smooth functions $\pi_j \widetilde{f}_N \quad \mathbf{R}^k \rightarrow \mathbf{R}$ through the data points $(\ X_i, \pi_j X_{i+1})$, $1 \leq i \leq N-1$, $j = 1, \ldots, k$ (π_j denotes the projection onto the ith- coordinate) $\widetilde{f}$ is , thus, a *predictor*

To measure the efficacy of $\widetilde{f}$ as a predictor, the normalized mean square error, or the *prediction error* is used as follows

$$\bar{E}^2 = \frac{\langle\ |\ X_{i+1} - \widetilde{f}\ (X_i)\ |^2\ \rangle}{\langle\ |\ X_i - \langle\ X_i\ \rangle\ |^2\ \rangle} \tag{4 1 3}$$

Thus, $\bar{E} = 0$ for perfect prediction and $\bar{E} = 1$ for predictions made using

$$\widetilde{X}_{i+1} = \langle\ X_n\ \rangle$$

4.2 Practical Considerations in Model Building

When a signal is measured over a period of time, it is assumed that the underlying dynamics of the system does not change. Considering the signal x_i, $i=1$, 2, .. to be the observed variable of the system, the dimension of the attractor is calculated. Usually, one calculates the correlation dimension D_2 because of ease of implementation. Next, the state-space reconstruction process involves the embedding of the observed signal in $\mathbf{R}^k$. That is, one must reconstruct the trajectory in a k-dimensional state-space from a single variable using delays as

given by equation (4 1 2) While Takens showed that $k = 2r+1$ is sufficient where r is the dimension of the manifold containing the attractor, a necessary condition for determinism is $k \geq D_2$ In practical situations, $k = 2r+1$ is rarely needed Moreover, if the dynamics can be described in a lower dimensional real space, it provides a lower model order description which is advantageous from the numerical computation viewpoint particularly in real time operations In equation (4 1 2), the delay between successive samples is equal to one In principle, this delay can be arbitrary In practice, however, if the delay is too small the coordinates become singular so that $X_j \simeq X_{j+1}$ If it is too big, then chaos makes X_1 and X_k causally disconnected A more systematic procedure based on mutual information was discussed in Chapter 3 The application of this method on all phoneme time series showed that most of them gave a minima of mutual information at a delay equal to one Hence, in all subsequent work we used this delay in reconstructing the state-space

4.3 Functional Representations

Once the state-space reconstruction process is complete, the next task is to fit a model to the data For the predictor $\tilde{f}$ to approximate chaotic dynamics as well, one must consider non-linear models From an infinite variety of such representations, an adhoc choice is made in the absence of greater theoretical understanding at the present juncture Descriptions of some such representations due to Farmer and Sidorowich [67] follow

1 Polynomials - They are the most widely used forms of representation partly because obtaining their parameters is a linear problem using least squares criterion Linear forms are used in auto-regressive and moving-average models while the more general m^{th}-degree non-linear polynomials are used in deterministic state-space modeling techniques, all of which have been reviewed in

Chapter 1 An mth- degree d- dimensional polynomial is

$$A_m (x_1, \quad ,x_d) = \sum_{l_1=0, \ldots, l_d=0} a_{l_1} \quad a_{l_d} \; x_1^{l_1} \quad x_d^{l_d} \qquad (4\ 3\ 1)$$

where $l_1+l_2+ \quad +l_d \leq m$ The number of parameters $a_{l_1} \quad a_{l_d}$ is $\frac{(d+m)!}{d!m!} \simeq d^m$ Fitting so many parameters for large m and d is impractical Also, polynomials have the disadvantage that they do not extrapolate well beyond their domain of validity

2 Rational Approximation - This is the ratio of two polynomials They extrapolate better than polynomials Also, fitting parameters by least squares criterion is linear

3 Radial Basis functions - They were recently suggested in the context of non-linear modeling by Casdagli [63] In their simplest form they depend only on the distance between points Thus,

$$R (x) = \sum_{n=1}^{N} \lambda_n \; \varphi (\| x - x_n \|), \qquad (4\ 3\ 2)$$

where $\| \; \|$ is the Euclidean norm and n is the label attached to points in the time series The coefficients λ_n are chosen to satisfy the interpolation conditions $x_{n+1} = R (x_n)$ A special case of radial basis functions are thin plate splines

4 Neural Networks - They are another class of functional representation A standard feed-forward neural net with two hidden layers can be written as

$$\hat{x}_{n+1} = \sum_k w_k \; z_k \; - a_0 \qquad (4\ 3\ 3)$$

$$z_k = \tanh (\sum_j w_j \; y_j \; - a_k) \qquad (4\ 3\ 4)$$

$$y_j = \tanh (\sum_l w_l \; x_l \; - a_j) \qquad (4\ 3\ 5)$$

where y_j and z_k are the values of the neurons in the two hidden layers and x_l are the input neurons The neurons are the coordinates in state-space A major disadvantage is that parameters cannot be fit by solving a linear problem As a result, fitting parameters takes several orders of magnitude more computer time

4.4 Global Approximation Technique

This is the first of the two approximation techniques to be discussed After the reconstruction of state-space, the chosen functional representation is used to approximate the scattered data points using one of the approximation techniques The global approximation technique uses all the data points in the state-space to get the approximating function $\tilde{f}$ The criterion generally used to get the parameters is the minimization of total squared error which reduces to a linear problem for polynomial and rational representations

Thus, given a block of observed data x_i, $i = 1, \quad , N$, one reconstructs the trajectory X_i, $i = 1, \quad , N-k+1$, in a k-dimensional state-space, where $k \geq D_2$, D_2 being the correlation dimension X_i is given by

$$X_i \quad = \quad [\; x_i \; x_{i+1} \; x_{i+2} \quad \quad x_{i+k-1} \;]^T \quad\quad\quad (4\ 4\ 1)$$

The global approximation technique using the least squares criterion is then a problem of finding the parameters of the approximant $\tilde{f}$, so that

$$\tilde{X}_{i+1} \quad = \quad \tilde{f} \; (\; X_i \;), \quad\quad 1 \leq i \leq N-k \quad\quad\quad (4\ 4\ 2)$$

and

$$E_j \quad = \sum_{i=2}^{N-k+1} | \; \pi_j X_i \; - \; \pi_j \tilde{X}_i \; |^2, \quad j = 1, \quad ,k \quad\quad\quad (4\ 4\ 3)$$

which is the total squared error for the jth- coordinate, is minimized for each coordinate j to obtain the parameters of the k approximating functions $\tilde{f}$

In our particular case, we have studied the performance of polynomial representations Moreover, we have used a delay m=1 for reconstructing the state-space as given by equation (4 4 1) Therefore, we have

$$\pi_j X_{i+1} \quad = \quad \pi_{j+1} X_i \; , \quad j = 1, \quad , k-1 \quad\quad\quad (4\ 4\ 4)$$

We preserve the above structure in the approximant $\tilde{f}$ This constraint is introduced to reduce the model order from the data compression point of view

Thus, for polynomial representations, we have

$$\pi_j \, \widetilde{X}_{i+1} \; = \; \pi_j \, \widetilde{f} \, (X_i) \; = \; \pi_{j+1} \, \widetilde{X}_i, \quad j = 1, \quad , k\text{-}1, \quad i=1, \quad , N\text{-}k \qquad (4\ 4\ 5)$$

$$\pi_k \, \widetilde{X}_{i+1} \; = \; \pi_k \, \widetilde{f} \, (X_i)$$

$$= \; c \; + \; \sum_{l=1}^{k} a_l \, \pi_l \, X_i \; + \; \sum_{l=1}^{k} \sum_{j=1}^{l} a_{lj} \, \pi_l X_j \, \pi_j X_i \qquad (4\ 4\ 6)$$

To obtain the parameters of the approximating function $\widetilde{f}$ using least squares criterion requires the solution of the following simultaneous linear equations

$$\sum_{i=1}^{N\text{-}k} \pi_k \widetilde{X}_{i+1} \; = \; \sum_{i=1}^{N\text{-}k} \pi_k X_{i+1} \qquad (4\ 4\ 7)$$

$$\sum_{i=1}^{N\text{-}k} \pi_k \widetilde{X}_{i+1} \, \pi_j X_i \; = \; \sum_{i=1}^{N\text{-}k} \pi_k X_{i+1} \, \pi_j X_i \, , \quad j = 1, \quad , k \qquad (4\ 4\ 8)$$

$$\sum_{i=1}^{N\text{-}k} \pi_k \widetilde{X}_{i+1} \, \pi_l X_i \, \pi_j X_i \; = \; \sum_{i=1}^{N\text{-}k} \pi_k X_{i+1} \, \pi_l X_i \, \pi_j X_i, \; j=1, \quad , k, \; l=1, \quad , j$$
$$(4\ 4\ 9)$$

where $\pi_k \widetilde{X}_{i+1}$ is given by equation (4 4 6)

The following section discusses the results of the global approximation technique as applied to speech signals in the form of phoneme utterances

4.5 Results and Discussion of the Global Approximation Technique

The global approximation technique using state-space reconstruction method discussed earlier provides us with a one-step predictor The performance of this prediction scheme was compared with the LPC The covariance method of the LPC employing a standard algorithm using Cholesky decomposition was used for comparison The comparison was in terms of the prediction error $\bar{E}^2$ given by equation (4 1 3), for the same model order p and blocklength N

For polynomial approximating functions, the model order p is obtained as follows Let the embedding dimension be k, i e , the trajectory X_i, $i = 1, 2, \quad , N\text{-}k$, in the reconstructed space is k-dimensional For the specific constraint

described in Section 4 4, which we have studied, the model orders for various degrees of approximating polynomial $\tilde{f}$ are given by

1 Linear $\qquad$ p = 1 + k,

2 Quadratic $\qquad$ p = 1 + k + k x (k + 1)/2,

3 Cubic $\qquad$ p = 1 + k + k x (k + 1)/2 + k x (k + 1) x (k + 2)/6,

etc

Thus, given the degree of the polynomial and the embedding dimension, the model order is fixed

The advantage of this modeling scheme over other schemes lies in the proper choice of the embedding dimension If the attractor dimension is D, then the choice of embedding dimension k is governed by ɪ) k $\geq$ D, k being an integer (necessary condition to preserve one-to-one relationship), ɪɪ) k $\geq$ 2r+1, where r is the dimension of the manifold containing the attractor (sufficient condition)

However, k $\geq$ 2r+1 is generally not required for most purposes For prediction purposes, one can increase k by 1 to get an idea of the variation of prediction error For phonemes which have D~ 2 0, the greatest drop in prediction error $\bar{E}^2$ was observed when k was changed from 1 to 2 (for representative examples, see Table 4 1) Thereafter, only minor drop in $\bar{E}^2$ was observed as k was increased even further For phonemes with larger values of D eg phoneme /z/, (Table 4 1 (c)), the drop in $\bar{E}^2$ continued until k ~ D This method has also been suggested for finding the required value of embedded dimension k

For purpose of reconstructing the state-space, the embedding dimension k was chosen as follows If D $<$ 3 0, then k=3 was used, else, k= D' was chosen where D' is the next higher integer after D For real time use of this modeling scheme, a common value of k chosen apriori is desirable This is because the calculation of attractor dimension for each block of data is not only time consuming but also

Embedding Dimension	Prediction Error	% Drop in Prediction Error from previous Embedding Dimension
1	2.00×10^{-2}	-
2	3.73×10^{-3}	436 %
3	2.52×10^{-3}	48 %
4	2.27×10^{-3}	11 %
5	2.17×10^{-3}	4.6 %
6	2.20×10^{-3}	- 1.36 %

(a)

Embedding Dimension	Prediction Error	% Drop in Prediction Error
1	8.88×10^{-2}	-
2	1.99×10^{-2}	346 %
3	1.81×10^{-2}	9.9 %
4	1.75×10^{-2}	3.4 %
5	1.59×10^{-2}	10.1 %

(b)

Embedding Dimension	Prediction Error	% Drop in Prediction Error
1	2.40×10^{-1}	-
2	1.81×10^{-1}	32.6 %
3	1.32×10^{-1}	37.1 %
4	9.80×10^{-2}	34.7 %
5	9.60×10^{-2}	2.1 %
6	9.25×10^{-2}	3.8 %
7	9.21×10^{-2}	- 0.44 %

(c)

Embedding Dimension	Prediction Error	% Drop in Prediction Error
1	4.70×10^{-2}	-
2	1.83×10^{-2}	157 %
3	1.73×10^{-2}	58 %
4	1.67×10^{-2}	36 %
5	1.60×10^{-2}	43 %

(d)

Table 4 1   The prediction error $\bar{E}^2$ with increasing embedding dimension k for blocklength N = 300 and quadratic polynomial for the following phonemes a) Phoneme /u/, Correlation Dimension D_2 = 2 00 ± 0 10, b) Phoneme /ɔ/, D_2 = 1 88 ± 0 05, c) Phoneme /z/, D_2 = 4 99 ± 0 31, d) Phoneme /k/, D_2 = 2 69 ± 0 39 The greatest drop in prediction error is when one changes from k=1 to k=2 for low D_2 For phonemes with higher D_2 eg c) the drop continues upto k ~ D_2 The advantage of this modeling approach is when one chooses the proper k (See text on the choice of embedding dimension)

partly manual (in the selection of the linear range of slopes in log C(r) vs log r

plots above the noise threshold) If $k > D$, a further increase of k leads to only a

slight drop in prediction error On the other hand, if $k < D$, an appreciable

decrease may occur If the degree of the approximating polynomial is m, then the

number of parameters required to be fitted is $\frac{(k+m)!}{k! \, m!}$ which depends heavily on k

and m Hence, a blanket choice of k greater than the attractor dimension of all

phonemes is costly in terms of computation As most phonemes have $D_2 < 4\,0$, a

compromise choice of k = 4 is suggested for speech modeling Thus, those regions

where $k > 4$, may not be approximated very well

In the comparison of the performance of global approximation technique

with the LPC (covariance method), blocklengths N = 50, 100, 200, 300, 400, 600 etc

were considered For global approximation, using quadratic polynomial and k=3

means parameter size p = 10 In all 10 phonemes were studied of which 5 were

vowels and 5 were consonants, including one each of voiced and unvoiced

fricatives and stops These phonemes are /u/, /U/, /o/, /ɔ/, /a/, /θ/, /ß/, /z/, /k/

and /g/ In all cases, both direct prediction (i e , where the blocklength used for

constructing the approximating polynomial equals the prediction length) and

iterative prediction (where blocklength for constructing the approximating

polynomial is less than the prediction length) were studied for different

blocklengths N

The general observations from the numerical computations are summarized

below (also see figures 4 1 and 4 2 for comparison of the error sequences due

the LPC and global approximation techniques) -

1 Choosing $N \geq 300$ and p=10, for vowels, LPC gave 2 to 10 times (more often 2 to

4 times) more prediction error than the global approximation technique (Table 4 2

gives representative examples)

Blocksize	Number of Blocks	Pred Error using LPC	Pred Error using global approx	% Drop in Predictior Error with global approx technique
50	27	$4\,94 \times 10^{-2}$	$2\,11 \times 10^{-3}$	2241
100	13	$1\,13 \times 10^{-2}$	$2\,44 \times 10^{-3}$	363
200	6	$1\,61 \times 10^{-2}$	$2\,48 \times 10^{-3}$	549
300	4	$7\,21 \times 10^{-2}$	$2\,51 \times 10^{-3}$	187

(a)

Blocksize	Number of Blocks	Pred Error using LPC	Pred Error using global approx	% Drop
50	16	$1\,93 \times 10^{-1}$	$9\,41 \times 10^{-3}$	1951
100	24	$3\,23 \times 10^{-2}$	$1\,24 \times 10^{-2}$	160
200	14	$2\,24 \times 10^{-2}$	$1\,32 \times 10^{-2}$	697
300	9	$1\,51 \times 10^{-2}$	$1\,44 \times 10^{-2}$	49
600	4	$1\,57 \times 10^{-2}$	$1\,48 \times 10^{-2}$	61

(b)

Blocksize	Number of Blocks	Pred Error using LPC	Pred Error using global approx	% Drop
50	24	$2\,47 \times 10^{-1}$	$5\,38 \times 10^{-2}$	359
100	12	$4\,64 \times 10^{-1}$	$9\,30 \times 10^{-2}$	399
200	6	$2\,02 \times 10^{-1}$	$1\,16 \times 10^{-1}$	741
300	4	$1\,50 \times 10^{-1}$	$1\,24 \times 10^{-1}$	210

(c)

Table 4 2 - Comparison of prediction error LPC against global approximation technique (quadratic polynomial) for specific cases   a) Phoneme /u/, Correlation Dimension $D_2 = 2\,00 \pm 0\,1$, Embedding Dimension k = 3 Thus, parametersize used , p = 10 b) Phoneme /k/, $D_2 = 2\,69 \pm 0\,39$, k = 3, p = 10/block, c) Phoneme /z/, $D_2 = 4.99 \pm 0.31$, k=5, p = 21/block

2 Choosing $50 \leq N \leq 100$, and $p = 10$ for vowels, LPC   gave 2 to 5 times more prediction error than the global approximation technique

3 Choosing $N \geq 200$ and $p = 10$, 20 etc , LPC gave 0 9 to 2 times the prediction error due to global approximation technique, for consonants

4 Choosing $50 \leq N \leq 100$ and $p = 10$, 20 etc LPC generally gave 2 to 10 times more prediction error than the global approximation technique, for consonants

5 For all phonemes, prediction error using linear polynomials was generally 1 02 to 2 0 times that using quadratic polynomials for the same blocks (see table 4 3 for the specific example of phoneme /s/)

6 We also compared the performance of the direct method against the iterative method The iterative method may be used when a lowering of the computational effort is needed The prediction error due to the iterative method is 1 02 to 2 0 times that due to the direct method As expected, for the same prediction length (eg $N_P = 600$) using successively increasing blocklengths for the approximating polynomial eg $(N_B = 50, 100, 200$ etc), the prediction error decreases Table 4 4 shows the prediction error using both direct and iterative methods for different blocklengths N_B and prediction lengths N_P for the phoneme /k/

Blocklength	Number of Blocks	Pred Error using Linear Polynomial	Pred Error using Quadratic Polynomial
50	55	$1 87 \times 10^{-2}$	$1 85 \times 10^{-2}$
100	29	$2 04 \times 10^{-2}$	$1 75 \times 10^{-2}$
200	14	$2 13 \times 10^{-2}$	$1 92 \times 10^{-2}$
300	9	$2 20 \times 10^{-2}$	$2 02 \times 10^{-2}$
600	5	$2 14 \times 10^{-2}$	$1 99 \times 10^{-2}$

Table 4 3 : Comparison of the prediction error using different blocklengths for the case of linear approximating polynomial vs quadratic polynomial for phoneme /s/, Correlation dimension $D_2 = 2 65 \pm 0 08$. Embedding dimension $k = 3$ was used Thus, parameter sizes for linear and quadratic polynomial approximations were $p=4$ and $p=10$ respectively

Pred Length	Pred Error using Direct method	Blocklength to construct approx polynomial	Pred Error using iterative method
50	$2\ 55 \times 10^{-2}$	50	-
100	$1\ 96 \times 10^{-2}$	50	$2\ 59 \times 10^{-2}$
200	$3\ 22 \times 10^{-2}$	50	$4\ 08 \times 10^{-2}$
300	$4\ 29 \times 10^{-2}$	50	$5\ 82 \times 10^{-2}$
400	$4\ 63 \times 10^{-2}$	50	$6\ 33 \times 10^{-2}$
100	$1\ 96 \times 10^{-2}$	100	-
200	$3\ 22 \times 10^{-2}$	100	$3\ 48 \times 10^{-2}$
300	$4\ 29 \times 10^{-2}$	100	$4\ 93 \times 10^{-2}$
400	$4\ 63 \times 10^{-2}$	100	$5\ 47 \times 10^{-2}$
200	$3\ 22 \times 10^{-2}$	200	-
300	$4\ 29 \times 10^{-2}$	200	$4\ 32 \times 10^{-2}$
400	$4\ 63 \times 10^{-2}$	200	$4\ 69 \times 10^{-2}$
500	$5\ 14 \times 10^{-2}$	200	$5\ 24 \times 10^{-2}$
600	$4\ 63 \times 10^{-2}$	200	$4\ 83 \times 10^{-2}$
300	$4\ 29 \times 10^{-2}$	300	-
400	$4\ 63 \times 10^{-2}$	300	$4\ 64 \times 10^{-2}$
500	$5\ 14 \times 10^{-2}$	300	$5\ 17 \times 10^{-2}$
600	$4\ 63 \times 10^{-2}$	300	$4\ 81 \times 10^{-2}$

Table 4 4 . Comparison of prediction of prediction error for the direct and iterative methods for different blocklengths N_B and prediction lengths N_P for the specific case of phoneme /k/, Correlation Dimension $D_2 = 2\ 69 \pm 0\ 39$ using quadratic approximating polynomials

Figure 4.1 . Comparison of error sequences using LPC prediction(Dashed line) and global approximation technique (Solid line) for phoneme /u/. Blocklength used= 100, parametersize = 10, prediction errors are 1.69×10^{-2} and $1 70 \times 10^{-3}$ respectively
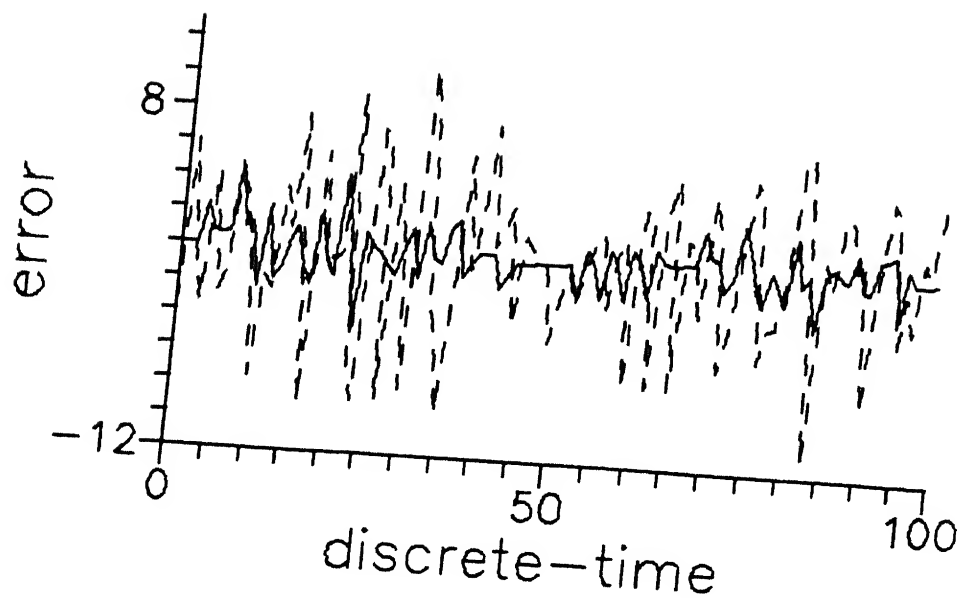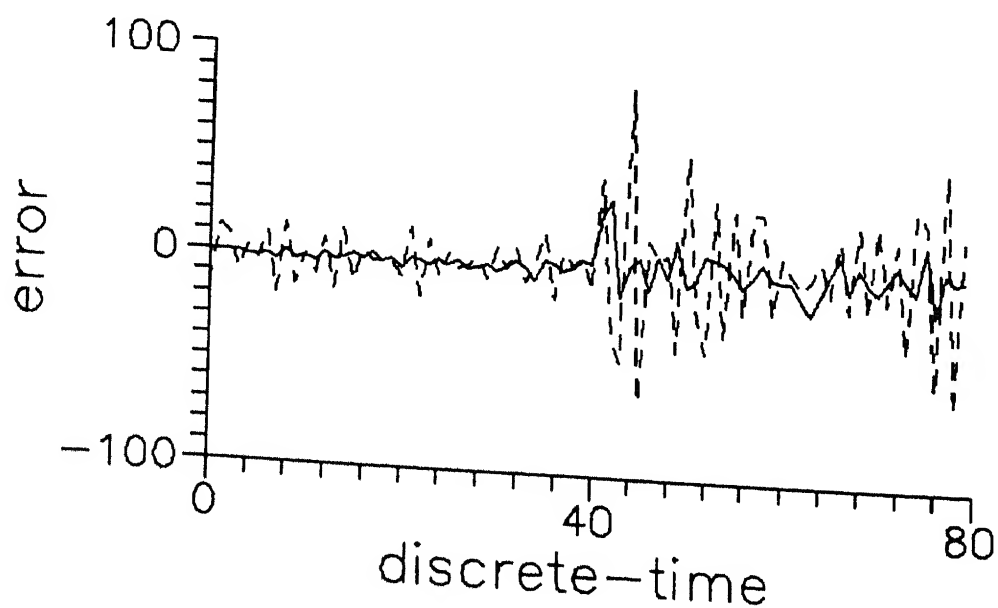

Figure 4.2 . Comparison of error sequences using LPC prediction (dashed line) and global approximation technique (Solid line) for phoneme /g/ Blocklength used= 80, parametersize = 10, prediction errors are $3 4 \times 10^{-2}$ and $2 90 \times 10^{-3}$ respectively.

The performance of the global approximation technique depends on the choice of the functional representation $\tilde{f}$ While many practical choices of functional representations exist, some of which were discussed in Section 4 3, our study was limited to the case of polynomial representation as a first step in this modeling scheme For a good approximation $\tilde{f}$ to the actual function f, $\tilde{f}$ must follow the variations of f The dependence on the choice of approximating function is reduced in the case of local approximation technique which will be described in the following section Also the better prediction properties of the global approximation technique over the LPC motivates us to study the local approximation technique which intuitively should give even better prediction properties

4.6 Local Approximation Technique

The basic idea here is to break up the state-space into local neighbourhoods and fit different parameters in each neighbourhood Thus, intuitively, local approximation should produce better fits for a given number of data points than global approximation, particularly for large blocklengths

The use of nearest neighbour approximation in the context of modeling chaotic dynamics was suggested by Farmer and Sidorowich [66,67] Their initial results show that global representations cannot be used to decrease the prediction error beyond a certain point even by adding more parameters or data Local approximation technique provides a means of using a chosen representation efficiently in the sense that beyond a certain point, adding more neighbourhoods gives more reduction in prediction error than going in for more parameters and using higher degree representation

There are three basic steps in the implementation of the local
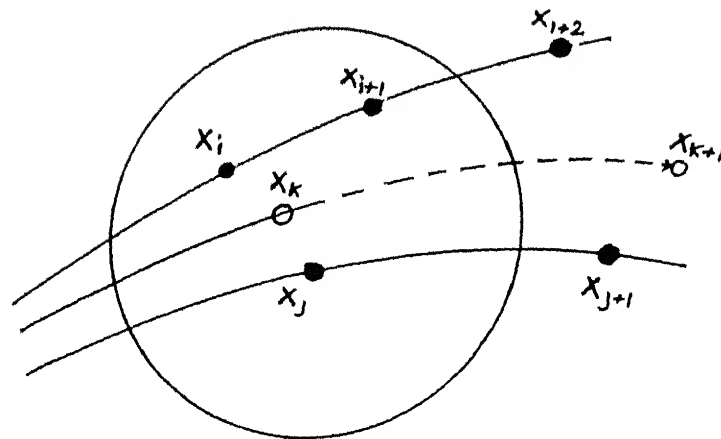
Figure 4.3 - Implementing local approximation technique -- Suppose one wants to predict the unknown state X_{k+1} from known X_k. Find the neighbourhood of X_k which in this case is denoted by black dots in the circle Choose a functional representation $\tilde{f}$ and fit the parametrs to these points using, for example, the least squares error criterion To make a prediction, evaluate $\tilde{f}$ at X_k

approximation technique in the neighbourhood of a point X in state-space

1 Pick a local functional representation,

2 Assign neighbourhoods,

3 Find a local chart that maps the points in each neighbourhood into their future value To make a prediction, evaluate the chart at X The basic idea is representated in figure 4 3

A simple way to assign neighbourhoods is to partition the domain into disjoint sets For example, one may use a rectangular grid Although this approach is convenient, it has the disadvantage that there is no overlap between the neighbourhoods and therefore no continuity between functions of one neighbourhood and another One way to overcome this problem is to introduce matching conditions between adjacent neighbourhoods This becomes a difficult

problem for data in more than two dimensions - a situation which arises very often in our case

An alternative that is more accurate than disjoint partitions and more convenient than implementing matching conditions is to overlap the neighbourhoods so that each function is constructed from a good set of neighbours A point Y_1 is said to be in the neighbourhood of point X if $| X-Y_1 | <$ R, R being a fixed radius

After getting the neighbourhood of point X, one fits the parameters of the chosen representation using the points in the neighbourhood For the case of polynomials, using the least square error criteria, the analysis is similar to the case of global approximation technique done in Section 4 5

Farmer and Sidorowich classify local approximation techniques according to the order of the derivatives the errors depend on Suppose that the function f being approximated is a polynomial of degree m If one wants to approximate f by a function $\widetilde{f}$ that is not itself a polynomial of degree m, then in the ideal case, the errors are proportional to ξ^{m+1}, where ξ is the spacing between the data points The average spacing between N points uniformly distributed over a D-dimensional space is $\xi \sim N^{-1/D}$ Calling q the *order of approximation*, then

$$\bar{E} \sim N^{-q/D} \tag{4 6 1}$$

where $\widetilde{E}$ is the r m s prediction error and in this case q = m+1

Farmer and Sidorowich then argue that acheiving the ideal case where q = m+1 is difficult for large q, since in general, fitting a polynomial of degree m does not produce a fit that is accurate to order m+1 Hence, they use equation (4 6 1) to define the order of local approximation, taking the limit as $N \rightarrow \infty$ and letting D be the information dimension

$$\alpha \quad = \quad - \lim_{N \to \infty} \frac{D \mid \log \bar{E} \mid}{\log N} \qquad (4\ 6\ 2)$$

In general, α depends on D, f, the way one chooses the neighbourhoods and other factors

Implementing the local approximation technique is more difficult compared to the global approximation technique If one uses disjoint neighbourhoods then the approximating functions have to be evaluated at each neighbourhood as opposed to one for global approximation Similarly, if one uses overlapping neighbourhoods, then the approximating function has to be evaluated at every point X_1 for predicting the next point X_{1+1} While this is acceptable for applications like weather forecasting where the sole aim is to obtain better predictions, such an approach is impractical for applications like data transmission etc where the model order must be kept under control In the following section we propose a compromised overlapping neighbourhood (CON)-local approximation technique and compare its prediction properties with global approximation

4.7 A Compromised Overlapping Neighbourhood (CON) – Local Approximation Technique

The proposed algorithm attempts to retain the better prediction properties of the local approximation technique and yet keeps the model order low for a fixed blocklength N The algorithm is similar to the local approximation technique except in the choice of neighbourhoods As seen in the previous section, the overlapping neighbourhood method is to find a set of points Y_1 to point X such that $\mid X - Y_1 \mid < R$, for fixed R The compromised overlapping method modifies this scheme in two respects Firstly, a certain minimum number of points NMINNEW are

included in each neighbourhood If for the chosen radius MRAD, the number of new

points NMRAD in the neighbourhood is less than NMINNEW, then (NMINNEW - NMRAD)

next nearest points are included in the neighbourhood Secondly, the same

functional representation is used for all points in the neighbourhood Thus, for a

given blocksize N, a maximum of [N/NMINNEW] functions are used for modeling the

block, setting an upper limit on the model order

The algorithm below gives the details of the implementation procedure with

discussions at appropriate places

The Algorithm

STEP 1   Read one block of data x_i, $i = 1,$ $, N$

STEP 2   Reconstruct the trajectory X_i, $i = 1,$ $, N-k+1$ in state-space of
dimension k

STEP 3   Read in constants MRAD and NMINNEW MRAD is the minimum radius of the
neighbourhood around a data point X_L NMINNEW is the minimum number of new
points to be included in the neighbourhood By *new* it is meant that any previous
neighbourhood did not contain the point

STEP 4   Initialize an array TALLY [1 N-k] with all 1's The entry 1 in a
location

TALLY [I] indicates that X_I has not yet been considered for prediction

STEP 5 i) $J = 0$,

       ii) $J = J + 1$, until TALLY [J] = 0,

J denotes the index of the 1st data point that has to be modeled for prediction

STEP 6 i) COUNTER = 0, L = J + 1,

      ii) COUNTER = COUNTER + 1,

      if TALLY [L] = 0 then

          L = L + 1 until TALLY [L] = 1

      SCHECK [COUNTER] = L,

      iii) if L = J + 1 then RCHECK = | X_J - X_L |, else

          RCHECK = max { | X_I - X_L |, I = SCHECK [1],   , SCHECK [COUNTER] }

iv) if RCHECK < MRAD, then

L = L + 1, goto ii)

else proceed to STEP 7.

STEP 6 finds a new point X_L which is approximately MRAD away from the first new point X_J in the neighbourhood. The new neighbourhood is created about the point X_L This step ensures that the maximum number of new points (i e , those not considered in any previous neighbourhood) are included See Figure 4 4

STEP 7 Find the neighbourhood $\{ X_i \}$ of point X_L as follows

All points $\{ X_i \}$ are in the neighbourhood of X_L such that

either $| X_L - X_i | <$ MRAD and NONEWPTS $\geq$ NMINNEW, where NONEWPTS is the number of new points in the neighbourhood,

or if NONEWPTS < NMINNEW, then

find (NMINNEW - NONEWPTS) new points $\{X_i\}$ which are the next nearest neighbours of X_L in the usual sense i e , TALLY [I] = 1 and

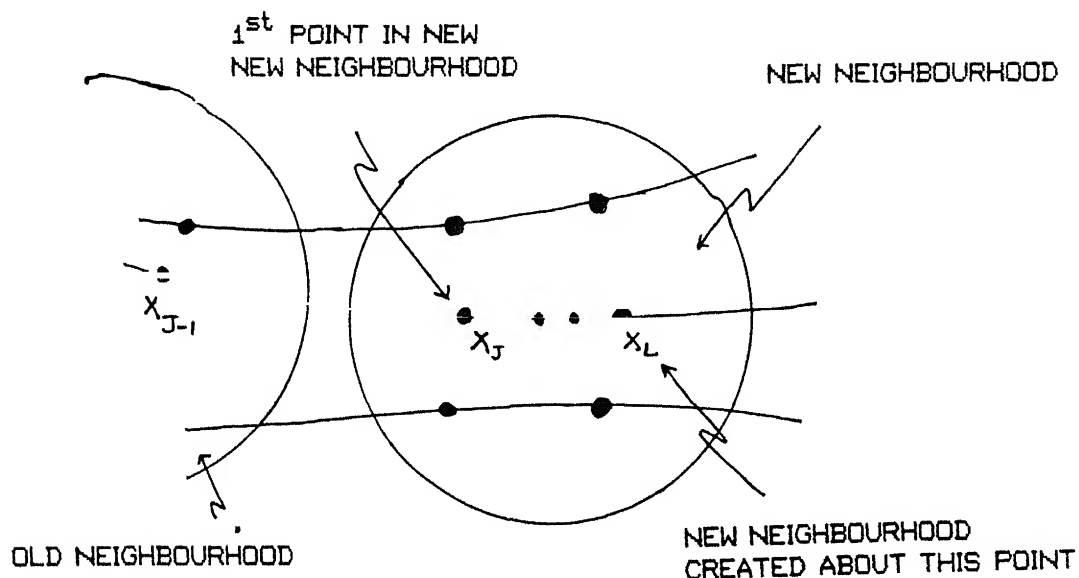$| X_L - X_i | < | X_L - X_j |$, for all $X_j \notin \{ X_i \}$

Figure 4 4   New neighbourhood created about point X_L and not X_J to include more new points in the neighbourhood This in turn helps in reducing the model order

STEP 8   ı) Store the neighbourhood value (X_1) and succeeding points (X_{1+1}) in matrix VAL

     ıı) TALLY [I] = 0 for all ı such that X_1 is in the neighbourhood

     Thus, at least NMINNEW points are in the new neighbourhood

STEP 9   Solve for the parameters of the approximating function $\tilde{f}$

STEP 10   Use $\tilde{f}$ to predict all new points $\tilde{X}_1$ in the neighbourhood

STEP 11   CHK = true,

     for COUNT = 1 to N-k do

       ıf TALLY [COUNT] = 1 then CHK = false,

    ıf CHK = false then goto STEP 5

     else fınd the error sequence and compute the prediction error

END OF ALGORITHM

     The prediction properties of this algorithm were compared with the global approximation technique The study was done on eight phonemes (4 vowels and 4 consonants) for two different blocklengths (N=600, 1000) at different locations in each time series For the global approximation technique, blocklengths N_G were considered such that the model order for both techniques remained the same This was achieved as follows The total number of neighbourhoods formed = NBD (or equivalently, the number of functions fitted to the entire blocklength) were noted in the CON-local approximation method Since the same functional representation was used in global approximation technique with the same embedding dimension k, blocklengths N_G used for the global approximation technique were given by N_G = [N/NBD] The prediction error was computed over NBD blocks Thus, the comparison of prediction error was for the same blocklength N and model order p The results based on the observations may be summarized as follows

     In 80 % of the cases studied, the CON - local approximation technique showed a further lowering of prediction error for the entire blocklength (N= 600 and 1000) over the global approximation technique This improvement was upto

over 100 % In the 20 % cases where the global approximation was better, the worst case was an increase in prediction error of 16 5 %

Table 4 5 shows the comparison for representative blocks of all the eight phonemes studied It is seen from the table that the improvement in the case of vowels is more pronounced than that for consonants Figures 4 5 and 4 6 show the error sequences for the CON - local approximation technique and the global approximation technique using identical blocklengths and model order

A comparison of the prediction properties thus shows that the global approximation technique is invariably better than the LPC while in 80 % cases the CON - local approximation technique is better than the global approximation technique in terms of prediction error This comparison is based on same blocklengths N and model order p Another important criterion in the choice of a modeling scheme is the order of computational complexity The following section discusses this aspect for the above modeling schemes

4.8 Computational Complexity Considerations

In the autocorrelation formulation of the LPC, as discussed in Chapter 1, the data outside a block of length N_1 is forced to zero for modeling the data within the block This leads to an error sequence that has large magnitude towards the ends of the block For this reason, the covariance method was used for the comparisons of prediction error with the global approximation technique However, the autocorrelation formulation leads to a Toeplitz matrix structure that can be more efficiently solved than the matrix structure obtained using the covariance method In the following, the computaitonal complexity of the autocorrelation and covariance methods of the LPC (see [79,81] for discussions) are compared with that of the global approximation technique Some discussion on the CON - local approximation technique is also given

S No	Pho-neme	Embedd Dimen	NMINNIW	Total no of nbds	Blocksiz	% improvement in pred error over global approx
1	/u/	3	50	10	600	20 6
2	/u/	3	100	9	600	16 4
3	/u/	3	50	5	1000	16 3
4	/u/	3	100	17	1000	15 6
5	/U/	3	50	11	600	18 0
6	/U/	3	100	5	600	28 2
7	/U/	3	50	18	1000	27 5
8	/U/	3	100	9	1000	44 3
9	/o/	3	50	11	600	47 9
10	/o/	3	100	5	600	14 9
11	/o/	3	50	17	1000	65 3
12	/o/	3	100	10	1000	31 2
13	/ɑ/	3	50	11	600	92 3
14	/ɑ/	3	100	5	600	103 0
15	/ʌ/	3	50	19	1000	113 0
16	/ʌ/	3	100	9	1000	86 7
17	/ɵ/	3	50	12	600	18 3
18	/ɵ/	3	100	6	600	3 2
19	/ɵ/	3	50	16	1000	51 9
20	/ɵ/	3	100	10	1000	22 0
21	/ʃ/	3	50	11	600	-10 5
22	/ʃ/	3	100	5	600	16 7
23	/ʃ/	3	50	14	1000	5 7
24	/ʃ/	3	100	9	1000	9 8
25	/z/	5	100	6	600	20 7
26	/z/	5	100	8	1000	- 1 7

Table 4 5   Representative examples of percentage decrease in the prediction error using the CON-local approximation technique over the global approximation technique for different phonemes

Figure 4 5   Comparison of error sequences using global approximation technique (dashed line) and CON-local approximation technique for phoneme /ʌ/, Blocklength = 600, MRAD = 12, NMINNEW=50 Prediction errors are 2.73×10^{-2} and 1.42×10^{-2} respectively, i e , an improvement of 92 3 %. Only part of the error sequence is shown
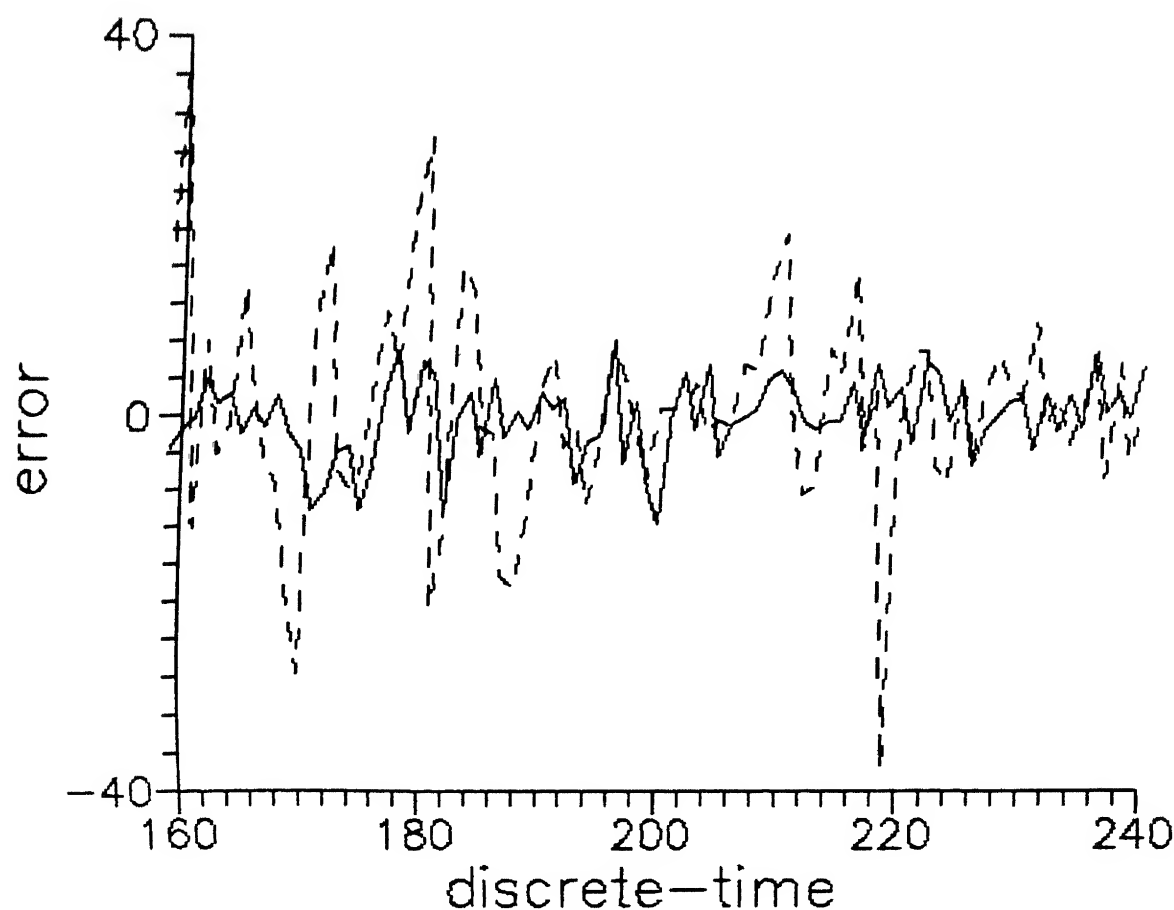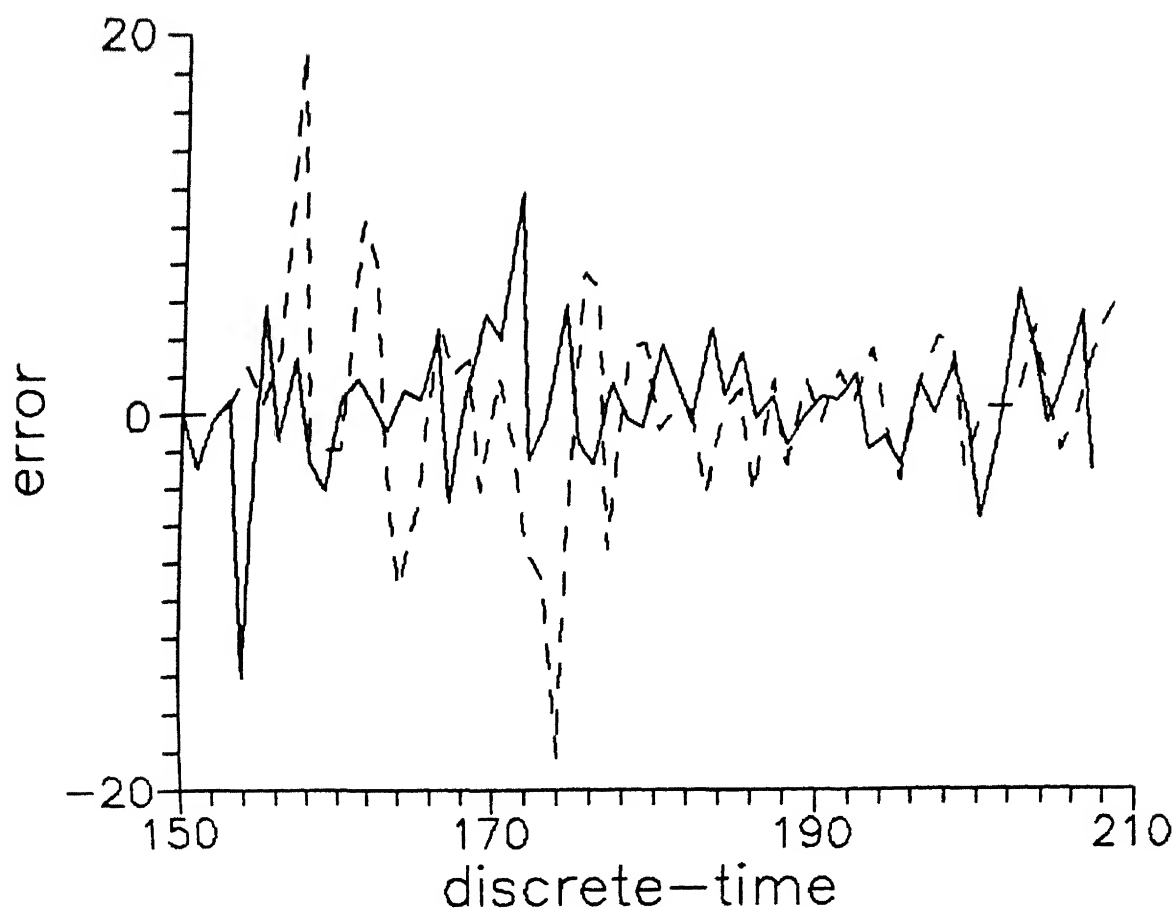
Figure 4.6 : Comparison of error sequences using global approximation technique (dashed line) and CON-local approximation technique for phoneme /o/. Blocklength = 600, MRAD = 12, NMINNEW = 50 Prediction errors are 8 09 x 10^{-3} and 5.47 x 10^{-3} respectively i e , an improvement of 47 9 %. Only part of the error sequence is shown

1 Autocorrelation Method of the LPC

The first step involves the multiplication of a block of data of length N_1 with a proper window function This requires N_1 multiplictions To obtain the p model parameters requires the solution of the following set of p simulatneous equations

$$\sum_{k=1}^{p} a_k \, R(\,|\,\imath - k\,|\,) = R(\imath), \quad 1 \leq \imath \leq p \qquad (4\ 8\ 1)$$

where

$$R(k) = \sum_{m=0}^{N-1-k} x(m) \; x(m+k) \qquad (4\ 8\ 2)$$

To obtain the coefficients of the correlation matrix, $O(N_1 p)$ multiplications are required Taking advantage of the Toeplitz structure of the matrix, one may use Durbin's method to solve equation $(4\ 8\ 1)$ This requires another $O(p^2)$ multiplications

2 Covariance Method of the LPC

Let the block of data be of length N_2 To model in terms of p coefficients requires the solution of

$$\sum_{k=1}^{p} b_k \, \varphi(\imath - k) = \varphi(\imath), \quad 1 \leq \imath \leq p \qquad (4\ 8\ 3)$$

where

$$\varphi(\imath - k) = \sum_{m=-k}^{N-k-1} x(m) \, x(m+k-\imath), \quad 1 \leq \imath \leq p, \, 0 \leq k \leq p \qquad (4\ 8\ 4)$$

Again, to obtain the matrix coefficients requires $O(N_2 p)$ multiplications The solution of the p simultaneous equations $(4\ 8\ 3)$ using Cholesky decomposition requires $O(p^3)$ multiplications

3 State-space Modeling using Global Approximation Technique

- The first step involves the reconstruction of the trajectory in a k-dimensional state-space This does not involve any multiplication process We have studied the special case of second degree polynomial representations The comparison with covariance method was in terms of same model order p for same blocklengths N_2 Using the least squares criteria led to the requirement of solving p simultaneous equations

(4 4 7) - (4 4 9) where,

$$p = 1 + k + k \times (k+1) / 2. \tag{4 8 5}$$

The next step is to find the entries of the coefficient matrix This requires $\left(\dfrac{N_2 - k}{4} \right)$ [$3k^4 + 10k^3 + 19k^2 + 12k$] multiplications which is O $(N_3 p^2)$ multiplications, where $N_3 = N_2 - k$, and p is given by equation (4 8 5) Finally, to solve the p simultaneous equations requires O (p^3) multiplications

4 State-Space Modeling using CON - Local Approximation Technique

The exact computational complexity in terms of the blocklength, parametersize etc cannot be given because of the dynamic nature of the algorithm Therefore, an idea of the computational complexity of the various stages of the algorithm is given

The first step of reconstructing the trajectory in a k-dimensional state-space does not involve any multiplication Thus, from a block of length N_2, a vector time series of length $N_3 = N_2 - k$ is created The most important step in this modeling scheme is to break the trajectory into local neighbourhoods The size of the local neighbourhoods depends upon the radius MRAD and the number of new points in the neighbourhood NMINNEW One solves equations (4 4 7)-(4 4 9) for each neighbourhood Hoewever, to obtain the coefficient matrices for all neighbourhoods requires O $(N_3 p^2)$ multiplications as in the global approximation case Finally, to solve the p simultaneous equations requires O (p^3) multiplications

The extra step involved in the CON - local approximation technique is to find m nearest neighbours to arrive at the neighbourhood Using brute force, this requires O (N_3) steps But using the efficient k-d algorithm [61,62], reduces this to O $(\log N_3)$ steps To obtain the m nearest neighbours requires O $(m\,2^k k)$ multiplications The value of m cannot be fixed apriori but it depends dynamically on MRAD and NMINNEW and the specific block of data

Type of Computation	Autocorrelation Method	Covariance Method	Global - Approximation Method
Windowing	N	0	0
Matrix Coefficients	$O(N_1 p)$	$O(N_2 p)$	$O(N_3 p)$
Matrix Solution	$O(p^2)$	$O(p^3)$	$O(p^3)$

Table 4 6 Computational considerations for the autocorrelation and covariance formulations of the LPC and the global approximation method of the deterministic state-space modeling $N_1 \approx N_2 \approx N_3$ denote the blocklength and p the parametersize

It is thus seen that as the prediction properties improve with succeeding models, the computational complexity increases. Table 4 6 summarizes the computational effort required in terms of multiplications at various stages of the autocorrelation, covariance and the global approximation methods

CHAPTER 5

CONCLUSION

A comprehensive theory has been developing for the past one and a half decades to describe a hitherto unexplained and yet an integral phenomenon of dynamical systems called *chaos* As emphasised at many places in the thesis, it is now known that apparently random behaviour can be due to the time evolution of simple non-linear systems Most of the initial efforts in chaos had been directed towards finding chaotic behaviour in specific systems This approach to analyse systems of differential equations governing the behaviour was based on first principles Quantities like Lyapunov exponents etc which describe chaotic behaviour, were computed analytically to check for chaos However, attention in the past two or three years has also been given to the modeling and prediction of chaotic behaviour Armed with the various tools developed to analyse chaotic behaviour eg dimensions and entropies, one can use them on observed scalar time series The attractor dimension gives an idea of the number of independent variables governing the system, while the entropy tells us whether the system is in the chaotic regime Even apparently random time series may have small dimensional attractors implying that it is the outcome of a deterministic system having few degrees of freedom Thus, viewing complexity as arising out of low dimensional chaos which is deterministic gives a new tool for analysing apparently random behaviour which is different from the theory of random processes One can use Takens' theorems to reconstruct the state-space from a single observed variable Thereafter, using simple deterministic non-linear modeling schemes, one can model observed behaviour

The thesis was initiated with an attempt to search for chaotic behaviour in speech signals This was, because many of the utterances eg fricatives, which are produced by a turbulent air flow through a narrow constriction, are complex to analyse and model using traditional techniques Other utterances eg vowels, are thought to be regular behaviour because of prominent frequencies called *formants* in their spectrum However, synthesised vowels based on the extracted formants sound artificial This led us to believe that although the formants are prominent, even in vowels the other frequenies cannot be ignored in any reconstruction process that attempts to retain the naturalness of sound

As pointed out in Chapter 3, a detailed model of the vocal tract should consider effects of time vari ation of the vocal tract shape, losses due to heat conduction and viscous friction at the vocal tract walls, softness of the vocal tract walls, radiation of sound at the lips, nasal coupling etc However, such a model has not yet been developed Consequently, our analysis of dimensions and entropy was based on phoneme utterances from the vocal tract instead of an analysis of any of the existing models from first principles Our analysis showed that phoneme time series are generated by low dimensional attractors Moreover, the analysis of the second-order entropy showed that most of the phoneme time series are chaotic in nature

The observation that the underlying attractors of phoneme time series are low-dimensional, allowed us to use the developing ideas of deterministic state-space modeling in this framework The results of the comparison of the global approximation technique using quadratic polynomials with the LPC were given in Chapter 4 For the same model order, the prediction error using global approximation technique was invariably lower than that of the covariance method of the LPC

The local approximation technique cannot be used directly for speech signal modeling in the context of storage and transmission because of the requirement of a prohibitively large model order Therefore, a compromised overlapping neighbourhood (CON) local approximation technique was proposed and its prediction properties compared with the global approximation technique It was observed that in 80 % of the cases studied, the CON - local approximation technique had lower prediction error than the global approximation technique

Success of a particular modeling scheme depends on the choice of the approximating function As a first step, quadratic polynomials were studied The global approximation technique is convenient to implement although its performance depends largely on the choice of the approximating functional representation This dependence can be intuitively seen to be reduced in the case of the CON-local approximation technique because modeling is limited to local areas However, a greater computational effort is required in this case Hence, the inherent problem of trade-off between performance and computational complexity remains The choice of a particular scheme depends upon the requirements and facilities available

Most of the work done in signal modeling in this thesis has been in the nature of gathering evidence using numerical procedures and computer simulation An analytical framework can only develop when the underlying theory stabilizes

Some suggestions for further work in this area are outlined below

1 Our work in the direction of ascertaining the chaotic nature of phoneme time series was limited to finding the second-order entropy A more fundamental way of ascertaining chaotic behaviour is to find the Lyapunov exponents for specific configurations of the system Finding all the Lyapunov exponents from a single

variable of observed data is a difficult task Nonetheless, some algorithms have been proposed recently [50,53,55] that can be used to get the largest Lyapunov exponents with a fair degree of confidence If the largest Lyapunov exponent is positive, it is a sufficient condition to establish the chaotic nature

2 Effects of quantizing the error sequence on the prediction properties can be studied and compared with the quantizers used with LPC models Since the state-space modeling schemes using global approximation and the CON-local approximation techniques can be used in ADPCM schemes for speech data transmission, a complete study of speech data compression aspects of suggested non-linear models of prediction using quantizers would be desirable Given the better prediction properties of the global and the CON-local approximation techniques, quantized error sequences are expected to give better reconstruction than the LPC which are currently the most popular models in ADPCM schemes Alternatively, lower number of bits may be required to transmit the error sequence using the new schemes to achieve the same level of reconstruction as with LPC models

3 The ability to model non-linear dynamics leads to a method for reducing external noise in the observations using non-linear averaging techniques [67] This scheme has been studied on data obtained from the evolution of differential equations State-space modeling techniques may be used to store speech data for future usage In such cases, non-linear averaging techniques may be used to improve the quality of sound Alternatively, speech signals generated in a noisy environment may be modeled using the above schemes and then non-linear averaging techniques may be applied to improve the quality of the reconstructed signal Given these uses, performance of non-linear averaging techniques applied to noisy speech signals can be studied

4 The modeling scheme studied in the thesis was limited to the polynomial class of functional representations As a next step, the class of rational functions may be studied Apart from the requirement of solving linear equations to get the parameters, they have the advantage of extrapolating better than polynomials Other classes eg neural network representations may also be studied

5 All the classic problems of speech processing can be studied in the framework of reconstructed state-space technique One can possibly do away with the artificial tool of time-warping employed in the problem of speech recognition

6 A lot of work is being done in applying Hidden Markov Models (HMM) to speech modeling The correspondence between the reconstructed state-space models and the hidden states of an HMM may also be studied in the above context of the nature of speech signals

Many complicated observed behaviour of systems in electrical engineering which were earlier explained by hand-waving arguments are now being analysed in the light of chaos Examples include non-linear circuits, phase-locked loops etc Another field that has opened up is the analysis and modeling of complex signals in the paradigm of chaos While we have studied some of these signal modeling aspects for the case of speech signals, they can also be used for other classes of signals eg biomedical signals, siesmic signals etc , some of which are now known to be chaotic These ideas can also be used to model image data for data compression applications

BIBLIOGRAPHY

The bibliography has been divided into 8 categories according to topics Some of the sources contained herein have not been referred to in the thesis, but have been included for the interest of the motivated reader Yet some others have not been consulted because of unavailability Nevertheless, they have been included because it is felt that they may be important for further work in this area

a. General Topics in Chaos and Fractals

1 H Atmanspacher and H Scheinberger, *Deterministic Chaos and Dynamical Instabilities in a Multimode CW Dye Laser*, Physical Review A, Vol 34, pp 253-263, 1986

2 M Barnsley, *Iterated Function Systems and Global Construction of Fractals*, Proc Royal Soc of London, Vol A399, pp 243-275, 1985

3 D Campbell, J P Crutchfield, J D Farmer and E Jen, *Experimental Mathematics The Role of Computation in Non-Linear Science*, Communications of the ACM, Vol 28, pp 374-384, 1985

4 J P Crutchfeld, J D Farmer, N H Packard and R S Shaw, *Chaos*, Scientific American, Vol 254, No 12, pp 46-57, 1986

5 J Gleick, Chaos, Cardinal, 1988

6 E N Lorenz, *Deterministic Non-Periodic Flow*, Journal of Atmospheric Science, Vol 20, pp 130-141, 1963

7 K Matshumoto and I Tsuda, *External Information in One-Dimensional Maps*, Physica, Vo 26D, pp 347-357, 1987

8 R May, *Simple Mathematical Models with very Complicated Dynamics*, Nature, Vol 261, pp 459-467, 1976

10 D Ruelle, *Strange Attractors*, Mathematical Intelligencer, Vol 2, pp 126-137, 1980

11 D Ruelle and F Takens, *On the Nature of Turbulence*, Comm of Mathematical Physics, Vol 20, pp 167-192, 1971

12 A M Saperstein, *Chaos - A Model for the Outbreak of War*, Nature, Vol 309,

pp 303-305, 1984

13 H L Swinney, *Observation of Order and Chaos in Non-Linear Systems*, Physica, Vol 7D, pp 3-15, 1983

14 D Triton, *Chaos in the Swing of a Pendulum*, New Scientist, Vol 111, pp 37-40, 24 July, 1986

b. Texts and Tutorials on Chaos and Fractals

15 M F Barnsley and S G Demko, Chaotic Dynamics and Fractals, Academic Press (N Y), 1985

16 R L Devaney, An Introduction to Chaotic Dynamical Systems, Benjamin/Cummings, 1986

17 J P Eckmann and D Ruelle, *Ergodic Theory of Chaos and Strange Attractors*, Reviews of Modern Physics, Vol 57, pg 617, 1985

18 N Gershenfeld, *An Experimentalist's Introduction to the Observation of Dynamical Systems*, Directions in Chaos, Vol 2, Ed H B Lin, World Scientific, pp 301-379, 1988

19 B Mandelbrot, Fractals Form, Chance and Dimension, Freeman (San Francisco), 1977

20 T S Parker and L O Chua, *Chaos   A Tutorial for Engineers*, Proc of the IEEE, Vol 75, No 8, pp 982-1008, 1987

21 D Ruelle, Chaotic Evolution and Strange Attractors, 1989

22 H G Schuster, *Deterministic Chaos*, Physik-Verlag, Weinheim, West Germany, 1984

c. Algorithmic Complexity and Randomness

23 G J Chaitin, *Randomness and Mathematical Proof*, Scientific American, Vol 232, No 5, pp 47-52, 1975

24 J Ford, *How Random is a Coin Toss ?*, Physics Today, Vol 36, No 4, pp 40-47, 1983

25 S Wolfram, *Origins of Randomness in Physical Systems*, Physical Review Letters, Vol 55, No 5, pp 449-452, 1985

d. Attractor Dimension and Entropy

26 A Ben-Mizrachi, I Procaccia and P Grassberger, *Characterization of Experimental (Noisy) Strange Attractors*, Physical Review A, Vol 29, No 2, pp 975-977, 1984

27 G Bennetin, L Galgani and J M Strelcyn, *Kolmogorov Entropy and Numerical Experiments*, Physical Review A, Vol 14, No 6, pp 2338-2345, 1976

28 A Brandstater, J Swift, H L Swinney, A Wolf, J D Farmer, E Jen and J P Crutchfield, *Low Dimensional Chaos in a Hydrodynamic System*, Physical Review Letters, Vol 51, No 16, pp 1442-1445, 1983

29 D S Broomhead and G P King, *Extracting Qualitative Dynamics from Experimental Data*, Physica, Vol 20D, pp 217-236, 1986

30 Elgar and M Kress, *Observations of the Fractal Dimension of Deep and Shallow Water Ocean Surface Gravity Waves*, Physica, Vol 37D, pp 104-108, 1989

31 J D Farmer, E Ott and J A Yorke, *The Dimension of Chaotic Attractors*, Physica, Vol 7D, pp 153-180, 1983

32 J D Farmer, *Information Dimension and Probabilistic Structure of Chaos*, Zeitschfigt Naturforschung, Vol 37 (a), pp 1304-1325, 1982

33 H Froehling, J P Crutchfield, D Farmer, N H Packard and R Shaw, *On Determining the Dimension of Chaotic Flows*, Physica, Vol 3D, pp 605-617, 1981

34 P Grassberger, *On the Hausdorff Dimension of Fractal Attractors*, Journal of Statistical Physics, Vo 26, No 1, pp 173-179, 1981

35 P Grassberger and I Procaccia, *Measuring the Strangeness of Strange Attractors*, Physica, Vol 9D, pp 189-208, 1983

36 P Grassberger and I Procaccia, *Characterization of Strange Attractors*, Physical Review Letters, Vol 50, No 5, pp 346-349, 1983

37 H S Greenside, A Wolf, J Swift, T Pignataro, *Impracticality of a Box-Counting Algorithm for Calculating the Dimensionality of Strange Attractors*, Physical Review A, Vol 25, No 6, pp 3453-3456, 1982

38 H G Hentschel and I Procaccia, *The Infinite Number of Generalized Dimensions of Fractals and Strange Attractors*, Physica, Vol 8D, pp 435-444, 1983

39 T Higuchi, *Approach to an Irregular Time-Series on the Basis of Fractal Theory*, Physica, Vol 31D, pp 277-283, 1988

40 Arun Kumar and S K Mullick, *On the Attractor Dimension and Entropy of Phonemes*, 1990 (to be published)

41 J Milner, *On the Concept of Attractor*, Comm of Mathematical Physics, Vol 99, pp 177-195, 1985

42 H Mori, *Fractal Dimensions of Chaotic Flows of Autonomous Dissipative Systems*, Progress of Theoretical Physics, Vol 63, No 3, 1980

43 N K Narayan and C S Sridhar, *Parametric Representation of Dynamical Instabilities and Deterministic Chaos in Speech*, Symposium on Signals, Systems and Sonars, Naval Physical and Oceanographic Laboratory, Cochin, India, pp B4 3/1-B4 3/4, 1988

44 E Ott and J D Hanson, *The Effect of Noise on the Structure of Strange Attractors*, Physics Letters A, Vol 85A, No 1, pp 20-22, 1981

45 I Procaccia, *The Static and Dynamic Invariants that Characterize Chaos and the Relations between them in Theory and Practice*, Physica Scripta, Vol T9, pp 40-46, 1985

46 A Rabinovitch and R Thieberger, *Time-Series Analysis of Chaotic Signals*, Physica, Vol 28D, pp 407-415, 1987

47 Roux, Simoyi and Swinney, *Observation of a Strange Attractor*, Physica, Vol 8D, pp 257-265, 1983

48 D A Russel, J D Hanson and E Ott, *Dimension of Strange Attractors*, Physical Review Letters, Vol 45, No 14, pp 1175-1178, 1980

e. Lyapunov Exponents

49 Benettin, Galgani, Giorgilli and Strelcyn, *Lyapunov Characteristic Exponents for Smooth Dynamical System and for Hamiltonian Systems   A Method to Compute all of them Part 1   Theory*, Meccanica, Vol 15, pg 9, 1980

50 J B Eckmann, S O Kamphorst, D Ruelle and S Ciliberto, *Liapunov Exponents from Time Series*, Physical Review A, Vol 34, No 6, pp 4971-4979, 1986

51 J Froyland, *Lyapunov Exponents for Multidimensional Orbits*, Physics Letters, Vol 97 A, No 1,2, pp 8-10, 1983

52 T Nagashima and I Shimada, *On the 'C-System' Like Property of the Lorenz System*, Progress in Theoretical Physics, Vol 58, No 4, pp 1318-1320, 1977

53 M Sano and Y Sawada, *Measurement of the Lyapunov Spectrum from a Chaotic Time Series*, Physical Review Letters, Vol 55, No 10, pp 1082-1085, 1985

54 I Shimada and T Nagashima, *A Numerical Approach to Ergodic Problem of Dissipative Dynamical Systems*, Progress in Theoretical Physics, Vol 61, No 6, pp

1605-1616, 1979

55 A Wolf, J B Swift, H L Swinney and J A Vastano, *Determining Lyapunov
Exponents from a Time-Series,* Physica, Vol 16D, pp 285-317, 1985

f. Reconstruction of Attractors

56 A M Fraser, *Information and Entropy in Strange Attractors,* IEEE Trans on
Information Theory, Vol 35, No 2, pp 245-262, 1989

57 A M Fraser, *Reconstructing Attractors from Scalar Time-Series   A Comparison
of Singular and Redundancy Criteria,* Physica, Vol 34D, pp 391-404, 1989

58 A M Fraser and H L Swinney,   *Independent Coordinates for Strange Attractors
from Mutual Information,* Physical Review A, Vol 33A, pp 1134-1140, 1986

59 N H Packard, J P Crutchfield, J D Farmer and R S Shaw, *Geometry from a
Time-Series,* Physical Review Letters, Vol 45, No 9, pp 712-716, 1980

60 F Takens, *Detecting Strange Attractors in Turbulence,* Dynamical Systems and
Turbulence, Warwick, 1980, Vol 898 Lect Notes in Mathematics, ed D A Rand and
L S Young, Springer, Berlin, pp 366-381, 1981

g. Deterministic State-Space Modeling

61   J H Bentley, *Multidimensional Binary Search Trees used for Associative
Searching,* Comm of the ACM, Vol 18, No 9, pp 509-517, 1975

62 J H Bentley, *Multidimensional Binary Search Trees in Database Applications,*
IEEE Trans of Software Engineering, Vol SE-5(4), pp 333-340, 1979

63 M Casdagli, *Non-Linear Prediction of Chaotic Time-Series,* Queen Mary College,
QMC Dyn 87 9, 1988 (preprint)

64 J Cremers and A Hubler,   *Construction of Differential Equations from
Experimental Data,* Z Naturforsch, Vol 42a, pp 797-802, 1987

65 J P Crutchfield and B S McNamara, *Equations of Motion from a Data Series,*
Complex Systems, Vol 1, pp 417-452, 1987

66 J D Farmer and J J Sidorowich, *Predicting Chaotic Time-Series,* Physical Review
Letters, Vol 59, No 8, pp 845-848, 1987

67 J D Farmer and J J Sidorowich, *Exploiting Chaos to Predict the Future and
Reduce Noise,* A Report, LA-UR-88-901, Los Alamos National Laboratory, 1988

68 J Guckenheimer, *Noise in Chaotic Systems,* Nature, Vol 298, pp 358-361, 1982

69 V Haggen, S M Heravi and M B Priestley, *A Study of the Application of State-Dependent Models in Non-Linear Time-Series Analysis,* Journal of Time-Series Analysis, Vol 5, No 2, pp 69-102, 1984

70 N Morrison, Introduction to Sequential Smoothing and Prediction, McGraw Hill Book Company, 1969

71 H Tong and K S Lim, *Threshold Autoregression, Limit Cycles and Cyclical Data,* Journal of the Royal Statistical Society B, Vol 42, No 3, pp 245-292, 1980

72 R Shaw, *Modeling Chaotic Systems,* Chaos and Order in Nature, Ed H Haken, Springer-Verlag, pp 218-231, 1981

h. Speech: Acoustic Theory and LPC Modeling

73 G Fant, Acoustic Theory of Speech Production, Mouton, The Hague, 1970

74 J L Flanagan, Speech Analysis, Synthesis and Perception, 2nd Edition, Springer Verlag, N Y 1972

75 J Makhoul, *Linear Prediction   A Tutorial Review,* Proc IEEE, Vol 63, pp 561-580, 1975

76 J Makhoul, *Speech Coding and Processing,* in Modern Signal Processing, Ed T Kailath, Hemisphere Publishing Corp , pp 211-247, 1985

77 J D Markel and A H Gray Jr , *Linear Prediction of Speech,* Springer-Verlag, N Y , 1976

78 M R Portnoff, *A Quasi-One Dimensional Digital Simulation for the Time-Varying Vocal Tract,* M S Thesis, Dept of E E , M I T , Cambridge, Mass , 1973

79 L R Rabiner and R W Schafer, *Digital Processing of Speech Signals,* Prentice Hall, 1978

80 M M Sondhi, *Model for Wave Propogation in a Lossy Vocal Tract,* Journal of Acoustic Society of America, Vol 55, No 5, pp 1070-1075, 1974

81 I H Witten, *Algorithms for Adaptive Linear Prediction,* The Computer Journal, Vol 23, No 1, pp 78-84, 1980

82 I H Witten, Principles of Computer Speech, Academic Press, London, 1982